



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁶ : A01N 37/18, 43/04, C12Q 1/00, 1/02, 1/68, C12N 5/00, 5/06, 15/00, 15/06, 15/09, 15/10, 15/11, G01N 33/53</p>	<p>A2</p>	<p>(11) International Publication Number: WO 98/54963 (43) International Publication Date: 10 December 1998 (10.12.98)</p>
<p>(21) International Application Number: PCT/US98/11422 (22) International Filing Date: 4 June 1998 (04.06.98) (30) Priority Data: 60/048,915 6 June 1997 (06.06.97) US 60/048,882 6 June 1997 (06.06.97) US (Continued on the following page) (71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): YOUNG, Paul [US/US]; 122 Beckwith Street, Gaithersburg, MD 20878 (US). GREENE, John, M. [US/US]; 872 Diamond Drive, Gaithersburg, MD 20878 (US). FERRIE, Ann, M. [US/US]; 13203 L Astoria Hill Court, Germantown, MD 20874 (US). RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). HU, Jing-Shan [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). OLSEN, Henrik, S. [DK/US]; 182 Kendrick Place #24, Gaithersburg, MD 20878 (US). EBNER, Reinhard [DE/US]; 9906 Shelburne Terrace #316,</p>		<p>Gaithersburg, MD 20878 (US). BREWER, Laurie, A. [US/US]; 14920 Mt. Nebo Road, Poolesville, MD 20837 (US). MOORE, Paul, A. [GB/US]; Apartment 104, 1908 Holly Ridge Drive, McLean, VA 22102 (US). SHI, Yanggu [CN/US]; 437 West Side Drive, Gaithersburg, MD 20878 (US). FLORENCE, Charles [US/US]; (US). FLORENCE, Kimberly [US/US]; 12805 Atlantic Avenue, Rockville, MD 20851 (US). LAFLEUR, David, W. [US/US]; 1615 Q Street, N.W. #807, Washington, DC 20009 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). FAN, Ping [CN/US]; Apartment 302, 335 West Side Drive, Gaithersburg, MD 20878 (US). WEI, Ying-Fei [CN/US]; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). FISCHER, Carrie, L. [US/US]; 5810 Hall Street, Burke, VA 22015 (US). SOPPET, Daniel, R. [US/US]; 15050 Stillfield Place, Centreville, VA 22020 (US). LI, Yi [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). ZENG, Zhizhen [CN/US]; 13950 Saddleview Drive, Gaithersburg, MD 20878 (US). KYAW, Hla [MM/US]; 520 Sugarbush Circle, Frederick, MD 21703 (US). YU, Guo-Liang [CN/US]; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). FENG, Ping [CN/US]; 4 Relda Court, Gaithersburg, MD 20878 (US). DILLON, Patrick, J. [US/US]; 1055 Snipe Court, Carlsbad, CA 92009 (US). ENDRESS, Gregory, A. [US/US]; 9729 Clagett Farm Drive, Potomac, MD 20854 (US). CARTER, Kenneth, C. [US/US]; 11601 Brandy Hall Lane, North Potomac, MD 20878 (US). (74) Agents: HOOVER, Kenley, K. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 10850 (US). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With declaration under Article 17(2)(a); without abstract; title not checked by the International Searching Authority.</i></p>
<p>(54) Title: 207 HUMAN SECRETED PROTEINS</p>		

(Continued)

60/048,892	6 June 1997 (06.06.97)	US	60/057,651	5 September 1997 (05.09.97)	US
60/048,901	6 June 1997 (06.06.97)	US	60/057,769	5 September 1997 (05.09.97)	US
60/048,900	6 June 1997 (06.06.97)	US	60/057,643	5 September 1997 (05.09.97)	US
60/048,893	6 June 1997 (06.06.97)	US	60/057,645	5 September 1997 (05.09.97)	US
60/048,964	6 June 1997 (06.06.97)	US	60/057,668	5 September 1997 (05.09.97)	US
60/048,884	6 June 1997 (06.06.97)	US	60/057,635	5 September 1997 (05.09.97)	US
60/048,894	6 June 1997 (06.06.97)	US	60/057,627	5 September 1997 (05.09.97)	US
60/048,971	6 June 1997 (06.06.97)	US	60/057,667	5 September 1997 (05.09.97)	US
60/048,885	6 June 1997 (06.06.97)	US	60/057,666	5 September 1997 (05.09.97)	US
60/049,375	6 June 1997 (06.06.97)	US	60/057,764	5 September 1997 (05.09.97)	US
60/048,881	6 June 1997 (06.06.97)	US	60/057,644	5 September 1997 (05.09.97)	US
60/048,880	6 June 1997 (06.06.97)	US	60/057,765	5 September 1997 (05.09.97)	US
60/048,896	6 June 1997 (06.06.97)	US	60/057,762	5 September 1997 (05.09.97)	US
60/049,020	6 June 1997 (06.06.97)	US	60/057,775	5 September 1997 (05.09.97)	US
60/048,876	6 June 1997 (06.06.97)	US	60/057,634	5 September 1997 (05.09.97)	US
60/048,895	6 June 1997 (06.06.97)	US	60/057,777	5 September 1997 (05.09.97)	US
60/049,019	6 June 1997 (06.06.97)	US	60/057,628	5 September 1997 (05.09.97)	US
60/048,916	6 June 1997 (06.06.97)	US	60/057,776	5 September 1997 (05.09.97)	US
60/048,970	6 June 1997 (06.06.97)	US	60/057,760	5 September 1997 (05.09.97)	US
60/048,972	6 June 1997 (06.06.97)	US	60/057,761	5 September 1997 (05.09.97)	US
60/048,949	6 June 1997 (06.06.97)	US	60/057,771	5 September 1997 (05.09.97)	US
60/048,974	6 June 1997 (06.06.97)	US	60/057,770	5 September 1997 (05.09.97)	US
60/048,883	6 June 1997 (06.06.97)	US	60/057,649	5 September 1997 (05.09.97)	US
60/048,897	6 June 1997 (06.06.97)	US	60/057,774	5 September 1997 (05.09.97)	US
60/048,898	6 June 1997 (06.06.97)	US	60/057,648	5 September 1997 (05.09.97)	US
60/049,373	6 June 1997 (06.06.97)	US	60/057,642	5 September 1997 (05.09.97)	US
60/048,917	6 June 1997 (06.06.97)	US	60/057,629	5 September 1997 (05.09.97)	US
60/048,962	6 June 1997 (06.06.97)	US	60/057,778	5 September 1997 (05.09.97)	US
60/048,878	6 June 1997 (06.06.97)	US	60/057,763	5 September 1997 (05.09.97)	US
60/049,374	6 June 1997 (06.06.97)	US	60/057,584	5 September 1997 (05.09.97)	US
60/048,875	6 June 1997 (06.06.97)	US	60/057,654	5 September 1997 (05.09.97)	US
60/048,899	6 June 1997 (06.06.97)	US	60/057,646	5 September 1997 (05.09.97)	US
60/048,877	6 June 1997 (06.06.97)	US	60/057,662	5 September 1997 (05.09.97)	US
60/048,963	6 June 1997 (06.06.97)	US	60/057,650	5 September 1997 (05.09.97)	US
			60/057,661	5 September 1997 (05.09.97)	US
			60/057,647	5 September 1997 (05.09.97)	US
			60/070,923	18 December 1997 (18.12.97)	US

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

207 Human Secreted Proteins

Field of the Invention

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and
5 their production.

Background of the Invention

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or
10 organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum
15 (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the
20 extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or
25 secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include
30 the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using
35 secreted proteins or the genes that encode them.

Summary of the Invention

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

Detailed Description

Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard,
5 Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained
10 in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42°C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the
15 filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages
20 of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH₂PO₄; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even
25 lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include
30 Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such
35 as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

5 The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and
10 double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability
15 or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

 The polypeptide of the present invention can be composed of amino acids joined
20 to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs,
25 as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be
30 branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a
35 nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

Polynucleotides and Polypeptides of the Invention

FEATURES OF PROTEIN ENCODED BY GENE NO: 1

This gene is expressed primarily in melanocytes and, to a lesser extent, in testes, ovary, kidney and other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer, disorders of neural crest derived cells including pigmentation defects, melanoma, reproductive organ defects, and defects of the kidney. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin,

reproductive, and renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating disorders that arise from alterations in the number or fate of neural crest derived cells including cancers such as melanoma and defects of the developing reproductive system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 2

This gene is expressed primarily in infant brain and fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental disorders of the brain or lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and pulmonary systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating or diagnosing disorders associated with abnormal proliferation of cells in the Central nervous system and developing lung.

FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene is expressed primarily in breast lymph node and to a lesser extent in ovarian cancer and chondrosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune responses such as inflammation or immune surveillance for

tumors. This gene may be important for inflammatory responses associated with tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 236 as residues: Lys-45 to Val-50, Lys-69 to Arg-76.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of immune responses including those associated with tumor-induced inflammation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 4

This gene is expressed primarily in T-cells and T-cell lymphomas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological diseases involving T-cells such as inflammation, autoimmunity, and cancers including T-cell lymphomas. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of T-cells and other cells of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating T-cell based disorders such as inflammatory diseases, autoimmune disease and tumors including T-cell lymphomas.

FEATURES OF PROTEIN ENCODED BY GENE NO: 5

This gene is expressed primarily in activated monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation, autoimmunity, infection, or disorders involving activation of monocytes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 238 as residues: Asp-19 to Arg-31.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing or treating diseases that result in activation of monocytes including infections, inflammatory responses or autoimmune diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 6

The translation product of this gene shares sequence homology with terminal deoxynucleotidyltransferase which is thought to be important in catalyzing the elongation of oligo- or polydeoxynucleotide chains.

This gene is expressed primarily in activated human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, particularly those of the blood such as leukemia and deficiencies in neutrophils such as neutropenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having

such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to terminal deoxynucleotidyltransferase indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and differential diagnosis of acute leukemia's. Alternatively, this gene may function in the proliferation of neutrophils and be useful as a treatment for neutropenia, for example, following neutropenia as a result of chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 7

The contig exhibits a reasonable homology to the human chorionic gonadotropic (HCG) analogue-GT beta-subunit as disclosed in U.S. Patent No. 5,508,261 and PCT Publication No. WO 92/22568. There is a high degree of conservation of the structurally important cysteine residues in these identities.

This gene is expressed primarily in IL-1 and LPS induced neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system, including inflammatory diseases and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases of the immune system since expression is primarily in neutrophils, and may be useful as a growth factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia following chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 8

This gene is expressed primarily in IL-1- and LPS-induced neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system, including inflammatory diseases and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 241 as residues: Ser-14 to Pro-22, Leu-43 to Val-53.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases of the immune system since expression is primarily in neutrophils, and may be useful as a growth factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia following chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 9

This gene is expressed primarily in IL-1 and LPS induced neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system, including inflammatory diseases and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 242 as residues: Tyr-22 to His-35.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases of the immune system since expression is primarily in neutrophils, and may be useful as a growth

factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia following chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 10

5 This gene is expressed primarily in activated T-cells and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
10 not limited to, immune dysfunctions including cancer of the T lymphocytes and autoimmune disorders and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at
15 significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of immune disorders particularly of T-cell origin and may act as a growth factor for particular subsets of T-cells such as CD4 positive cells which would make this a useful therapeutic for the treatment of HIV and other immune compromising illnesses.

FEATURES OF PROTEIN ENCODED BY GENE NO: 11

This gene is expressed primarily in fetal tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
30 biological sample and for diagnosis of many developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing fetus, expression of this gene at significantly higher or lower levels may be routinely detected
35 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor or differentiation factor for particular cell types in the developing fetus and may be useful in replacement or other types of therapy in cases where the gene is expressed aberrantly.

FEATURES OF PROTEIN ENCODED BY GENE NO: 12

This gene is expressed primarily in T-cells and to a lesser extent in tumor tissue including glioblastoma, meningioma, and Wilm's tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system including autoimmune conditions such as rheumatoid arthritis, inflammatory disorders and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 245 as residues: Thr-9 to Ser-14.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis/ modulation of immune function disorders, including rheumatoid arthritis and inflammatory responses.

FEATURES OF PROTEIN ENCODED BY GENE NO: 13

This gene is expressed primarily in placenta and to a lesser extent in fetal liver and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of hematological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of

disorders of the above tissues or cells, particularly of the hematological and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
 5 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor for hematopoietic stem cells or
 10 progenitor cells in the treatment of chemotherapy patients or kidney disease.

FEATURES OF PROTEIN ENCODED BY GENE NO: 14

This gene is expressed primarily in stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as
 15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of hematopoietic disorders including cancer, neutropenia, anemia, and thrombocytopenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of
 20 the above tissues or cells, particularly of the hematopoietic and immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,
 25 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor for hematopoietic stem cells or progenitor cells, in particular following chemotherapy treatment.

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 15

The translation product of this gene shares sequence homology with epsilon-COP from *Bos taurus* which is thought to be important as a component of coatamer, a complex of seven proteins, that is the major component of the non-clathrin membrane coat. Preferred polypeptides encoded by this gene comprise the following amino acid
 35 sequences:

MAPPAPGPASGGSGEVDELFDVKNAFYIGSYQQCINEAXXVKLSSPERDVERD

VFLYRAYLAQRKFGVVLDEIKPSSAPELQAVRMFADYLAHESRRDSIVAELDRE
 MSRSXDVTNTTFLMAASIYLHDQNPDAALRALHQGDSLECTAMTVQILLKLD
 RLDLARKELKRMQDLDEDATLTQLATAWVSLATGGEKLQDAYYIFQEMADKCS
 PTTTTLLNGQAACHMAQGRWEAAEGLLQEALDKDSGYPETLVNLIVLSQHLGKP
 5 PEVTNRYLSQLKDAHRSHPIKEYQAKENDFDRLVLQYAPSAEAGPELSGP
 (SEQ ID NO:458); or RDVERDVFLYRAYLAQRKFGVVLDEIKPSSAPELQAVRMF
 ADYLAHESRRDSIVAELDREMSRSXDVTNTTFLMAASIYLHDQNPDAALRALH
 QGDSLECTAMTVQILLKLDRLDLARKELKRMQDLDEDATLTQLATAWVSLATG
 GEKLQDAYYIFQEMADKCSPTLLLLNGQAACHMAQGRWEAAEGLLQEALDKD
 10 SGYPETLVNLIVLSQHLGKPPEVTNRYLSQLKDAHRSHPIKEYQAKENDFDRL
 VLQYAPSA (SEQ ID NO:459).

This gene is expressed primarily in activated monocytes and T-cells, and to a lesser extent in multiple other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as
 15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunomodulation, specifically relating to transport problems in these cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell
 20 type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene
 25 expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to epsilon-COP indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating /diagnosing problems with the cellular transport of proteins that may result in
 30 immunologic dysfunction.

FEATURES OF PROTEIN ENCODED BY GENE NO: 16

The translation product of this gene shares sequence homology with an RNA
 helicase which is thought to be important in polynucleotide metabolism. The translation
 35 product of this contig exhibits good homology to the LbeIF4A antigen of *Leishmania braziliensis*. The LbeIF4A antigen, or immunogenic portions of it, can be used to induce protective immunity against leishmaniasis, specifically *L. donovani*, *L. chagasi*,

L. infantum, *L. major*, *L. braziliensis*, *L. panamensis*, *L. tropica* and *L. guyanensis*. It can also be used diagnostically to detect *Leishmania* infection or to stimulate a cellular and/or humoral immune response or to stimulate the production of interleukin-12.

This gene is expressed primarily in colon cancer and to a lesser extent in
5 pituitary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of cancers particularly of the colon. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
10 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the gastrointestinal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
15 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 249 as residues: Glu-93 to Ala-98, Gln-150 to Leu-156, Leu-220 to Leu-231, Leu-268 to Arg-273, Val-324 to Pro-341, Arg-372 to Asn-380, Ser-405 to Gly-410, Phe-426 to Ala-433, Glu-458 to Asp-470, Arg-506 to Ser-547.
20

The tissue distribution and homology to RNA helicase indicates that polynucleotides and polypeptides corresponding to this gene are useful for development of diagnostic tests for colon cancer.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 17

The translation product of this contig has sequence homology to a cytoplasmic protein that binds specifically to JNK designated the JNK interacting protein-1 or JIP-1 in mice. JIP-1 caused cytoplasmic retention of JNK and inhibition of JNK-regulated
30 gene expression.

This gene is expressed primarily in brain including pituitary cerebellum frontal cortex, fetal brain and to a lesser extent in the kidney cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
35 biological sample and for diagnosis of the central nervous system disorders including ischemia, epilepsy, Parkinson's disease, and schizophrenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological

probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Furthermore, the translation product of this contig may suppress the effects of the JNK signaling pathway on cellular proliferation, including transformation by the Bcr-Abl oncogene. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 250 as residues: Pro-6 to Ser-26, Ala-30 to Asp-41, Gly-55 to Ser-61, Gly-74 to Thr-80, Tyr-117 to Ala-123, Tyr-167 to Asp-172, Ala-212 to Cys-223, Pro-239 to Tyr-244.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for enhanced survival and/or differentiation of neurons as a treatment for neurodegenerative disease.

FEATURES OF PROTEIN ENCODED BY GENE NO: 18

The translation product of this gene shares sequence homology with a liver stage antigen from a protozoan parasite.

This gene is expressed primarily in fetal tissue and to a lesser extent in activated T-cells and other immune cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities and diseases of immune function. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to a protozoan antigen indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/immune modulation of parasitic infections.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 19

Preferred polypeptide encoded by this gene comprise the following polypeptide sequences:

MKAIGIEPSLATYHHIIRLFDQPGDPLKRSSFIIYDIMNELMGKRFS PKD
 PDDDKFFQSAMSICSSLRDLELAYQVHGLLKTGDNWKFIGPDQHRNFYYSKFF
 10 DLICLMEQIDVTLKWYEDLIPSA YFPHSQTMHLLQALDVANRLEVIPKIWER
 (SEQ ID NO:460); and/or KDSKEYGHTFRSDLREEILMLMARDKHPPELQVAF
 ADCAADIKSAYESQPIRQTAQDWPATSLNCIALFLRAGRTQEAWKMLGLFRKH
 NKIPRSELLNELMDSAKVSNSPSQAIEVVELASAFSLPICEGLTQRVMSDFAINQ
 EQKEALSNLTALTSDSDTDSSSDSDSDTSEGK (SEQ ID NO:461). Polynucleotides

15 encoding such polypeptides are also provided.

This gene is expressed primarily in stromal and CD34 depleted bone marrow cells and to a lesser extent in tissues of embryonic origin.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
 20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of hematologic origin including cancers and immune dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of
 25 the hematopoietic and immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily
 30 fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 252 as residues: Ser-28 to Gln-34.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor for hematopoietic stem cells or progenitor cells which may be useful in the treatment of chemotherapy patients
 35 suffering from neutropenia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 20

Preferred polypeptide fragments can be found in an alternative open reading frame. These preferred polypeptides comprise the amino acid sequence:

MSSDNESDIEDLKLRLRDLKHLKEIQDLQSRQKHEIESLYTKLGKVPPAVI
 5 IPPAAPLSGRRRRPTKSKGSKSSRSSLGNKSPQLSGNLSGQSAASVLHPQQTL
 HPPGNIPESGQNQLLQPLKSPSSDNLYSFTSDGAISVPSLSAPGQGTSTNTV
 GATVNSQAAQAQPPAMTSSRKGTFTDDLHKLVDNWARDAMNLSGRRGSKGH
 MNYEGPGMARKFSAPGQLCISMTSNLGGAPISAASATSLGHFTKSMCPPQQY
 GFPATPFGAQWSGTGGPAPQLGQFQPVGTASLQNFNISNLQKSISNPPGSNL
 10 RTT (SEQ ID NO:462); IQDLQSRQKHEIESLYTKLGKVPPAVIIPPAAPLSGRRRR
 PTKSKGSKSSRSSLGNKSPQLSGNLSGQSAASVLHPQQTLHPPGNIPESGQN
 QLLQPLKSPSSDNLYSFTSDGAISVPSLSAPGQGTSTT (SEQ ID NO:463);
 TSDGAISVPSLSAPGQGTSTNTV GATVNSQAAQAQPPAMTSSRKGTFTDDLH
 (SEQ ID NO:464); KGHMNYEGPGMARKFSAPGQLCISMTSNLGGAPISAAS
 15 ATSLGHFTK (SEQ ID NO:465); QPLKSPSSDNLYSFTSDGAISVPSLSAPG
 (SEQ ID NO:466). Also preferred are polynucleotide fragments encoding these
 polypeptide fragments.

This gene is expressed in fetal liver and tissues associated with the CNS.

Therefore, polynucleotides and polypeptides of the invention are useful as
 20 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, liver and CNS diseases. Similarly, polypeptides and antibodies directed
 to these polypeptides are useful in providing immunological probes for differential
 identification of the tissue(s) or cell type(s). For a number of disorders of the above
 25 tissues or cells, particularly of the liver and CNS, expression of this gene at
 significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
 fluid or spinal fluid) or another tissue or cell sample taken from an individual having
 such a disorder, relative to the standard gene expression level, i.e., the expression level
 30 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred
 epitopes include those comprising a sequence shown in SEQ ID NO: 253 as residues:
 Gln-26 to Lys-34.

The tissue distribution indicates that polynucleotides and polypeptides
 corresponding to this gene are useful for diagnosis and treatment for liver diseases such
 35 as hepatocellular carcinomas and diseases of the CNS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 21

In an alternative reading frame, this gene shows sequence homology to two recently cloned genes, karyopherin beta 3 and Ran_GTP binding protein 5. (See Accession Nos. gil2102696 and gnlIPIDle328731.) The Ran_GTP binding protein is related to importin-beta, the key mediator of nuclear localization signal (NLS)-dependent nuclear transport. Based on homology, it is likely that this gene may activity similar to the RAN_GTP binding protein. Preferred polypeptide fragments comprise the amino acid sequence: VRVAAAESMXLLLECA XVRGPEYLTQMWHFMCDALIKA IGTEPDS DVLSEIMHSFAK (SEQ ID NO:467). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed in thymus tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 22

This gene is expressed primarily in prostate and osteoclastoma tissues. Preferred polypeptide fragments also comprise the amino acid sequence: MEINNQNCFIVIDLVRTVMENGVEGLLIFGAFLPESWLIGVRCSSPEPKALLLIL AHSQKRRLDGWSFIRHLRVHYCVSLTIHFS (SEQ ID NO:468). Also preferred are polynucleotide sequences encoding this polypeptide fragment.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, bone and prostate diseases, and cancers, particularly of the bone and prostate. Similarly, polypeptides and antibodies directed to these polypeptides are

useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone and prostate systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded

5 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 255 as residues: Met-1 to Ser-11.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for bone and prostate disorders, especially cancers of those systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 23

15 This gene shares sequence homology with the FK506-binding protein (FKBP-13) family, a known cytosolic receptor for the immunosuppressants. Recently, another group has cloned a very similar gene, recognizing the homology to FK506-binding protein family, calling their gene FKBP23. (See Accession No. 2827255.)

This gene is expressed primarily in lymphoid tissues.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample, especially for those susceptible to immune suppressant therapies and for diagnosis of diseases and conditions, which include, but are not limited to, immune suppressant disorders. Similarly, polypeptides and antibodies directed to these

25 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

30 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 256 as residues: Ala-19 to Val-31, Arg-38 to Gly-49, Ala-61 to Lys-66, Tyr-68 to Pro-78, Gly-116 to Ala-121, Asp-154 to

35 Ser-162, Glu-173 to Gln-186, Phe-194 to Gly-203, Pro-207 to Val-212.

The tissue distribution and homology to FKBP-12 and -13 indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for immune suppressant disorders.

5 **FEATURES OF PROTEIN ENCODED BY GENE NO: 24**

This gene is expressed primarily in the brain and in the retina. This gene maps to chromosome 8, and therefore can be used in linkage analysis as a marker for chromosome 8.

Therefore, polynucleotides and polypeptides of the invention are useful as
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological and ocular associated disease states. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of
15 disorders of the above tissues or cells, particularly of the disorders of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard
20 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 257 as residues: Cys-34 to Asp-40.

The tissue distribution in retina indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and/or detection of eye disorders
25 including blindness, color blindness, impaired vision, short and long sightedness, retinitis pigmentosa, retinitis proliferans, and retinoblastoma. Expression in the brain indicates a role in the is useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive
30 disorder and panic disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 25

This gene shows sequence homology to a newly identified class of proteins expressed in the nervous system, called stathmin family. (See Accession No. 2585991;
35 see also Eur. J. Biochem. 248 (3), 794-806 (1997).) The stathmin family appears to be an ubiquitous phosphoprotein involved as a relay integrating various intracellular signaling pathways. These pathways affect cell proliferation and differentiation.

Preferred polypeptide fragments comprise the amino acid sequence:

QDKHAEVRKNKELKEEASR (SEQ ID NO:469); QQDLSPWAAPVGCPLXXASX
TCHXLPLSGCLRRQSXSLPVVAXLCFWFSCPLASLFVPGQPCVTCFPFSLPFQD
KHAEVRKNKELKEEASR (SEQ ID NO:470). Also preferred are the

5 polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
10 not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to
these polypeptides are useful in providing immunological probes for differential
identification of the tissue(s) or cell type(s). For a number of disorders of the above
tissues or cells, particularly of the central nervous system, expression of this gene at
significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
15 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
fluid or spinal fluid) or another tissue or cell sample taken from an individual having
such a disorder, relative to the standard gene expression level, i.e., the expression level
in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
20 corresponding to this gene are useful for the detection/treatment of neurodegenerative
disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's
Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive
compulsive disorder and panic disorder.

25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 26**

The polynucleotide sequence of this gene contains a domain similar to a Flt3
ligand peptide. Preferred polypeptide fragments comprise the amino acid sequence:
PTRCCTTQPCRSSARRPCWVPMVPSPEGREXQPTCPS (SEQ ID NO:471). Thus,
this gene may have activity as binding to Flt3 receptors, a process known to promote
30 angiogenesis and/or lymphangiogenesis.

This gene is expressed in human tonsil, and to a lesser extent in
teratocarcinoma, placenta, colon carcinoma, and fetal kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for identification of the tissue(s) or cell type(s) present in a biological sample
35 and for diagnosis of diseases and conditions, which include, but are not limited to,
diseases of the tonsil, as well as cancers, such as colon, reproductive, and kidney
cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful

in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tonsils, colon, reproductive organs, and kidneys, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 259 as residues: Pro-22 to Glu-33.

The tissue distribution in tonsil and several cancers and fetal tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the tonsil or colon, such as tonsillitis, inflammatory diseases involving nose and paranasal sinuses, especially during the infection of influenza, adenoviruses, parainfluenza, rhinoviruses. The gene may also be useful in the diagnosis and treatment of neoplasms of nasopharynx or colon origins.

FEATURES OF PROTEIN ENCODED BY GENE NO: 27

In an alternative reading frame exists a large open reading frame that encodes a preferred polypeptide. Preferred polypeptide fragments comprise the amino acid sequence:

MKRSLNENSARSTAGCLPVPLFNQKKRNRQPLTSNPLKDDSGISTPSDNYDFP
 PLPTDWAWEAVNPEXAPVMKTVDTGQIPHSVSRPLRSQDSVFNSIQSNTGRSQ
 GGWSYRDGNKNTSLKTWXKNDFKPQCKRTNLVANDGKNSCPMSSGAQQQK
 QLRTPEPPNLSRNKETELLRQTHSSKISGCTMRGLDKNSALQTLKPNFQQNQY
 KXQMLDDIPEDNTLKETSLYQLQFKEKASSLRISA VIESMKYWREHAQKTVLL
 FEVLAVLDSAVTPGPYYSKTFLMRDGKNTLPCVFYEIDRELRLIRGRVHRCVG
 NYDQKKNIFQCVSVRPASVSEQKTFQAFVKIADVEMQYYINVMNET (SEQ ID
 NO:472); SQDSVFNSIQSNTGRSQGGWSYRDGNKNTSLKTWXKNDFKPQCKR
 (SEQ ID NO:473); NKETELLRQTHSSKISGCTMRGLDKNSALQTLKPNF (SEQ ID
 NO:474); SSLRISA VIESMKYWREHAQKTVLLFEVLAVLDSAVTPGPYYSKTFLM
 (SEQ ID NO:475); and PRLIRGRVHRCVGN YDQKKNIFQCVSVRPASVSEQKT
 FQAFV (SEQ ID NO:476).

This gene is expressed primarily in human testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, male reproductive disorders, including cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a hormone with reproductive or other systemic functions; contraceptive development; male infertility of testicular causes, such as Klinefelter's syndrome, varicocele, orchitis; male sexual dysfunctions; testicular neoplasms; and inflammatory disorders such as epididymitis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 28

This gene is expressed primarily in apoptotic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases relating to T cells, as well as cancer in general. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the disorders of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune disorders. Moreover, since the gene was isolated from an apoptotic cell and based on the understanding of the relationship of apoptosis and cancer, it is likely that this gene may play a role in the genesis of cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 29

This gene is expressed primarily in human tonsils.

Therefore, polynucleotides and polypeptides of the invention are useful as
 5 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, gastrointestinal disorders. Similarly, polypeptides and antibodies directed
 to these polypeptides are useful in providing immunological probes for differential
 identification of the tissue(s) or cell type(s). For a number of disorders of the above
 10 tissues or cells, particularly of the gastrointestinal system, expression of this gene at
 significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
 fluid or spinal fluid) or another tissue or cell sample taken from an individual having
 such a disorder, relative to the standard gene expression level, i.e., the expression level
 15 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
 corresponding to this gene are useful for the diagnosis and treatment of gastrointestinal
 diseases.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 30

The translation product of this gene shares sequence homology with C44C1.2
 gene product of *Caenorhabditis elegans* with unknown function. Preferred polypeptide
 fragments comprise the amino acid sequence:

GVFRPCVCGRPASLTCSPLDPEVGOPYCDTPTMRTLFNLLWLALACSPVHTTLSK
 25 SDAKKAASKTLEKSQFSDKPVDGRGLVVTDLKAESVVLEHRSYCSAKARDRH
 FAGDVLGYVTPWNSHGYDVTKVFGSKFTQISPVWLQLKRRGREMFVETGLHD
 VDQGWMRVVRKHAKGLHIVPRLLFEDWTYDDFRNVLDSEDEIEELSKTVVQVA
 KNQHFDGFVVEVWNQLLSQKRVGLIHMLTHLAEALHQARLLALLVIPAITPGT
 DQLGMFTHKEFEQLAPVLDGFSMTYDYSTAHQPGPNAPLSWVRACVQVLDP
 30 KXKWRTKSSWGSTSMXWTXRPXDARXPVVGXRXIQXLKDHXPRMVLDISK
 PQ (SEQ ID NO:477); TCSPLDPEVGOPYCDTPTMRTLFNLLWLALACSPVHTTLS
 (SEQ ID NO:478); LVVTDLKAESVVLEHRSYCSAKARDRHFAGDVLGYVTPW
 NSHGYDVTKVFGSKF (SEQ ID NO:479); REMFEVTGLHDVDQGWMRVVRK
 HAKGLHIVPRLLFEDWTYDDFRNVLDSEDE (SEQ ID NO:480); HFDGFVVEVW
 35 NQLLSQKRVGLIHMLTHLAEALHQARLLALLVIPAITPGTDQLGM (SEQ ID
 NO:481); DGFSLMTYDYSTAHQPGPNAPLSWVRACVQVLDPKXKWRTKSSW
 GST (SEQ ID NO:482). Also preferred are polynucleotide fragments encoding these

polypeptide fragments. This gene maps to human chromosome 11, and therefore is useful in linkage analysis as a marker for chromosome 11.

This gene is expressed primarily in human T cells and to a lesser extent in human colon carcinoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and gastrointestinal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 263 as residues: Leu-21 to Ala-30, Ser-38 to Asp-47, Pro-87 to Asp-94, Leu-197 to Thr-204, Pro-256 to Ser-262, Thr-277 to Arg-282, Thr-293 to Trp-303.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders and gastrointestinal diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 31

The translation product of this gene shares sequence homology with Ribosomal protein L11 of *Caenorhabditis elegans*. (See Accession No. 156201.) Preferred polypeptide fragments comprise the amino acid sequence:

ERGVSINQFCKEFNERTKDIKEGIPLPTKILVKPDRTFEIKIGQPTVSYFLKAAAG
IEKGARQTGKEVAGLVTLKHVYEIARIKAQDEAFALQDVPLSSVVRISIIGSARSL
GIRVVKDLSSSEELAAF QKERAIFLAAQKEADLAAQEAAKK (SEQ ID NO:483).

Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed in human embryo tissue and to a lesser extent in human epithelioid sarcoma and other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, development disorders and epithelial cell cancer. Similarly, polypeptides and antibodies

directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryonic and epithelial cell systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 264 as residues: Lys-34 to Gly-40.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of developmental disorders and epithelial cancer.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 32**

This gene is expressed primarily in resting T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory and general immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders of immune system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 33

This gene is believed to reside on chromosome 1. Accordingly, polynucleotides derived from this gene are useful in linkage analysis as chromosome 1 markers.

This gene is expressed primarily in prostate and to a lesser extent in soares adult brain, human umbilical vein endothelial cells, and amniotic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, prostate-related disorders. Similarly, polypeptides and antibodies
 5 directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the urinary system and nervous system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
 10 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful
 15 for the diagnosis and treatment of disorders of the urinary and nervous systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 34

This gene shares sequence homology with R05G6.4 gene product. (See Accession No. gil1326338.) This gene also shares sequence homology with the cyclophilin-like protein
 20 CyP-60. (See Accession No. 1199598, see also Biochem. J. 314 (1), 313-319 (1996).) Preferred polypeptide fragments comprise the amino acid sequence:
 AVYTYHEKKKDTAASGYGTQNIRLSRDAVKDFDCCCLSLQPCHDPVVTPDGYL
 YERE AILEYILHQKKEIARQMKA YEKQRGTRREEQKELQRAASQDHVRGFLEKE
 SAIVSRP LNPFTAKALSGTSPDDVQPGPSVGPPSKDKDKVLPSFWIPSLTPEAK
 25 ATKLEKPSRTVTCMSGKPLRMSDLTPVHFTPLDSSVDRVGLITRSEYVCAVT
 RDSLSNATPCAVLRPSGAVVTLECEKLRKDMVDPVTGDKLTDRDIIVLQRT
 (SEQ ID NO:484); YLYERE AILEYILHQKKEIARQMKA YEKQRGTRREEQKELQ
 RAASQDHVRGFLE (SEQ ID NO:485); and FTAKALSGTSPDDVQPGPSVGPP
 SKDKDKVLPSFWIPSLTPEAKATKLEKPSRTVTCMSGKPL (SEQ ID NO:486).
 30 Also preferred are polynucleotide fragments that encode these polypeptide fragments.

This gene is expressed primarily in human testis and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
 35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male reproductive disorders and in particular testicular cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system. Expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders of the male reproductive system and in particular of testicular cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 35

The translation product of this gene shares sequence homology with Lpe5p of *Saccharomyces cerevisiae* which is thought to be important in the metabolism of phospholipids.

This gene is expressed primarily in liver and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic and nervous systems expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 268 as residues: Pro-14 to Leu-20, Lys-28 to Asn-38, Arg-109 to Arg-114, Lys-119 to Asn-124, Glu-152 to Leu-157, Pro-172 to Val-180.

The tissue distribution and homology to Lpe5p of *Saccharomyces cerevisiae* indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of metabolic and nervous disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 36

This gene shares sequence homology with the nuclear ribonucleoprotein U (HNRNP U), encoded by *C. elegans* (See Accession gil1703576.) Preferred polypeptide fragments comprise the amino acid sequence:

- 5 MDTSENRPENDVPEPPMPIADQVSNDDEEGSVEDEEKKESSLPKSFKRKISVV
SATKGVPAAGNSDTEGGQPGRKRRWGASTATTQKKPSISITTESLKSLIPDIKPL
AGQEAVVDLHADDSETERNGDDGTHDKGLKICRTVTQVVPAGQENGQ
REEEEEKEPEAEPPVPPQVSVEVALPPPAEHEVKKVTLGDTLTRRSISQKSGV
SITIDDPVRTAQVSPPRGKISNIVHISNLVRPFTLGQLKELLGRTGTLVEEAFWI
10 DKKSHCFVTYSTVEEAVATRTALHGVKWPQSNPKFLCADYAEQDEL DYHRGL
LVDRPSETKTEEQGI PRPLHPPPPPVQPPQHPRAEQREQERAVREQWAERERE
MERRERTRSEREWD RDKVREGPRSRSRSRXRRRKERAKSKEKKSEKKEKAQE
EPPAKLLDDLFRKTKAAPCTIYWLPLTDSQIVQKEAERAERAKEREKRRKEQEEE
EQKEREKEAERERNRQLEREKRREHSRERDRERERERERDRGDRDRDRERDRE
15 RGRERDRRDTKRHSRSRSTPVRDRGGR (SEQ ID NO:488). Also preferred are
the polynucleotide fragments encoding this polypeptide fragments.

This gene is expressed primarily in epididymus.

- Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, diseases of the male reproductive system. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the male reproductive system, expression of
25 this gene at significantly higher or lower levels may be routinely detected in certain
tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
30 disorder.

The tissue distribution indicates that polynucleotides and polypeptides
corresponding to this gene are useful for the diagnosis and treatment of male
reproductive disorders.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 37

This gene is expressed primarily in amygdala.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory diseases and reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the amygdala, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of inflammatory diseases and reproductive disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 38

This gene shares sequence homology with human opsonin protein P35 fragment. (See Accession No. R94181.) The opsonin protein activates the phagocytosis of pathogenic microbes by phagocytic cells. Preferred polypeptide fragments comprise the amino acid sequence: GCDSCPHLPREAFQAQDTQAEGECSSRAERADMCPDAP PSQEVPEGPGAAP (SEQ ID NO:489). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed in immune-related tissues such as thymus, macrophage, T cells and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders and infectious disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and infectious disease, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 271 as residues: Lys-9 to Arg-14, Met-38 to Asp-51.

- 5 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders, as well as the treatment and/or diagnosis of infectious disease.

FEATURES OF PROTEIN ENCODED BY GENE NO: 39

- 10 The translation product of this gene shares sequence homology with alpha-2 type I collagen which is thought to be important in tissue repair. (See, e.g., 211607.) Preferred polypeptide fragments comprise the amino acid sequence: PQLPSCGRPW PGTASVFQSHTQGPREDPDPCRAQGSAGTHCPISLSPPRQ (SEQ ID NO:490). Also preferred are the polynucleotide sequences encoding these polypeptide sequences.

- 15 This gene is expressed primarily in the brain and to a lesser extent in the kidney and thymus

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain, kidney, and immune disorders. Similarly, polypeptides and
20 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, kidney, and immune disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
25 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 30 The tissue distribution and homology to alpha-2 type I collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tissue repair, and brain, kidney, immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 40

- 35 The translation product of this gene shares sequence homology with mini-collagen which is thought to be important in tissue repair tumor metastasis. (See Accession No. gnl|PID1006976.) Preferred polypeptide fragments comprise the amino acid sequence: PGFRGPSGLGCSFFPRSLGRVLPPGCQRPGAHAD

SSPPPTP (SEQ ID NO:491). Also preferred are polynucleotides encoding this polypeptide fragment.

This gene is expressed in ovarian cancer and to a lesser extent in dendritic cells and smooth muscle.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumor metastasis and tissue repair. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes
10 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumor metastasis and tissue repair, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
15 an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 273 as residues: Asn-2 to His-11.

The tissue distribution and homology to mini-collagen gene indicates that
20 polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tumor metastasis and tissue repair.

FEATURES OF PROTEIN ENCODED BY GENE NO: 41

This gene shares sequence homology with the HIV TAT protein. (See
25 Accession No. 328416.) Preferred polypeptide fragments comprise the amino acid sequence: EDLKKPD PASLRAASCGEGKKRKACKNCTCGLAE ELEKEK SREQMSSQPKSACGNCYLGD AFRASC PYLGMPAFKPGEKVLLS (SEQ ID NO:492); EDLKKPD PASLRAASCGEGKKRKACKNCTCGLAE ELEKEK SREQMSSQPKSACGNCYLGD AFRASC PYLGMPAFKPGEKVLLS SDSNLHD
30 (SEQ ID NO:493); CGNCYLGD AFRASC PYLGMPAFKPGEKVLLS SDS (SEQ ID NO:494); SCGEGKKRKACKNCTCGLAE ELEKE (SEQ ID NO:495); SQPKSAC GNCYLGD AFRASC (SEQ ID NO:496); and REAGQNSERQYVS LSRD (SEQ ID NO:497). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

35 This gene is expressed primarily in the infant brain and to a lesser extent in the breast and testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain, testes and breast disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, testes and breast disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 274 as residues: Pro-7 to Val-15.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of brain, testes and breast, and other related disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 42

This gene is expressed primarily in the infant brain, human cerebellum, and to a lesser extent in medulloblastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain related disorders and medulloblastoma and other brain cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain related disorders and brain cancers, including medulloblastoma, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 275 as residues: Thr-41 to Glu-47.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of human brain related disorders, brain cancers, and medulloblastoma.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 43

The translation product of this gene shares sequence homology with a phosphotyrosine-independent ligand for the lck SH2 domain which is thought to be important in signal transduction related to phosphotyrosine-independent ligand for the lck SH2 domain. (See Accession No. gil1184951.) Preferred polypeptide fragments
10 comprise the amino acid sequence: ESSGQARTLADPGPGWPRQQGMCFGSLT
GLSTTPHGFLT VSAEADPRLIESLSQMLSMGFSDEGGWLTRLLQTKNYDIGAAL
DTIQYSKH (SEQ ID NO:498). Also preferred are polynucleotide fragments encoding this polypeptide fragment. It is likely that this gene is a new member of a family of
phosphotyrosine-independent ligands for the lck SH2 domains.

15 This gene is expressed primarily in the placenta and to a lesser extent in endothelial cells and neutrophil.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
20 not limited to, reproductive, cardiovascular, immune, and infectious diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular, reproductive, and immune system, and infectious diseases, expression
25 of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the
30 disorder.

The tissue distribution and homology to a phosphotyrosine-independent ligand for the lck SH2 domain indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cardiovascular, reproductive, and immune system diseases, as well as infectious diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 44

This gene is expressed primarily in the fetal brain, cerebellum and to a lesser extent in the placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, neuronal cell related disorders. Similarly, polypeptides and antibodies
directed to these polypeptides are useful in providing immunological probes for
differential identification of the tissue(s) or cell type(s). For a number of disorders of
10 the above tissues or cells, particularly of the neuronal cell related disorders, expression
of this gene at significantly higher or lower levels may be routinely detected in certain
tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level, i.e.,
15 the expression level in healthy tissue or bodily fluid from an individual not having the
disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
NO: 277 as residues: Thr-20 to Gly-28.

The tissue distribution and homology to proline-rich protein genes indicates that
polynucleotides and polypeptides corresponding to this gene are useful for diagnosis
20 and treatment of neuronal cell related disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 45

The translation product of this gene shares sequence homology with
precerebellin of human, which is thought to be important in synaptic physiology. (See
25 Accession No. gil180251.) It has been observed that cerebellin-like immunoreactivity is
associated with Purkinje cell postsynaptic structures. Thus, it is likely that this gene
also have synaptic activity. Preferred polypeptide fragments comprise the amino acid
sequence: QEGSEPVLLEGECLVVCEPGRAAAGGPGGAALGEAPPGRVAFXAV
RSHHHEPAGETGNGTSGAIYFDQVLVNEG GGFDRASGSFVAPVRGVYSFRFH
30 VVKVYNRQTVQVSLMLNTWPVISAFANDPDVTREAAATSSVLLPLDPGDRVSLR
LRRGXSTGW (SEQ ID NO:499). Also preferred are polynucleotide fragments
encoding these polypeptide fragments.

This gene is expressed primarily in cerebellum and infant brain. By Northern
analysis, a single transcript of 2.4 kb was observed in brain tissues.

35 Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, neuronal cell signal transduction and synaptic physiology. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal cell signal transduction and synaptic physiology expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to gene or gene family indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal cell related disorders.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 46**

This gene is expressed in fetal liver and spleen, and to a lesser extent in bone marrow, umbilical vein, and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the immune system, particularly hematopoiesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoiesis and immune disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 279 as residues: Asp-30 to Glu-57.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hematopoietic and immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 47

The translation product of this gene shares sequence homology with a 12 kD nucleic acid binding protein of Feline calicivirus which is thought to be important in viral replication. (See Accession No. 59264)

5 This gene is expressed primarily in human cardiomyopathy and to a lesser extent in T helper cells, fetal brain and synovial sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiomyopathy as well as viral infection. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 280 as residues: Trp-20 to Cys-26.

The tissue distribution in cardiomyopathy and homology to viral 12 kD nucleic acid binding protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of cardiomyopathy, including those caused by ischemic, hypertensive, congenital, valvular, or pericardial abnormalities.

25 The gene expression pattern may be the consequence or the cause for these conditions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 48

The translation product of this gene shares sequence homology with tumor necrosis factor related gene product which is thought to be important in tumor necrosis, bacterial and viral infection, immune diseases and immunoreactions.

This gene is expressed primarily in colon and to a lesser extent in ovarian and breast cancers.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors of colon, ovary or breast origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the colon, ovary and breast, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to Tumor necrosis factors indicates that polynucleotides and polypeptides corresponding to this gene are useful for intervention of cancers of colon, ovary and breast origins, because TNF family members are known to be involved in the tumor development.

FEATURES OF PROTEIN ENCODED BY GENE NO: 49

The translation product of this gene shares sequence homology with mucins, such as epithelial mucin, which is thought to be important in extracellular matrix functions such as protection, lubrication and cell adhesion (See for example Accession No. R68002). Preferred polypeptide fragments comprise the following amino acid sequence: PRSRPALRPGRQRPPSHSATSGVLRPRKKPDP (SEQ ID NO:500).

Also preferred are polynucleotide fragments encoding these polypeptide fragments. Moreover, this gene maps to chromosome 22q11.2-qter, and therefore, can be used as a marker in linkage analysis for chromosome 22.

This gene is expressed primarily in corpus colosum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors, especially of corpus colosum, as well as metastatic lesions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the corpus colosum and other solid tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to mucins indicates that polynucleotides and polypeptides corresponding to this gene are useful for serum tumor markers or immunotherapy targets because tumor cells have greatly elevated level of mucin expression and shed the molecules into the epithelial tissues.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 50

This gene is expressed primarily in CD34 depleted buffy coat cord blood and primary dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic disorders and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in CD34 depleted buffy coat cord blood and primary dendritic cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hematopoietic and immune disorders. Secreted or cell surface proteins in the above tissue distribution often are involved in cell activation (e.g. cytokines) or molecules involved in cell surface activation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 51

The translation product of this gene shares sequence homology with Interferon induced 1-8 gene encoded polypeptide which is thought to be important in binding to retroviral rev responsive element. Preferred polypeptide fragment comprise the following amino acid sequences: MTLITPSXKLTFXKGNKSWSSRACSSSTLVDP (SEQ ID NO:501). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in CD34 positive cells and neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, retroviral infection, such as AIDS, and other immune disorders.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 284 as residues: Gln-51 to Trp-62.

The tissue distribution and homology to interferon induced gene 1-8 indicates that polynucleotides and polypeptides corresponding to this gene are useful for intervention of retroviral infection including HIV. The factor may be involved in viral stability or viral entry into the cells. Alternatively, the virus/factor complex may elicit the cellular immune reaction.

FEATURES OF PROTEIN ENCODED BY GENE NO: 52

This gene shares sequence homology to immunoglobulin lambda chain (See Accession No. 2865484). Therefore it is likely that this gene has activity similar to an immunoglobulin lambda chain. Preferred polypeptide fragments comprise the following amino acid sequence: GHPSPALSIAPSDGSQPCDEVYPYGEAHVTRYCKKPLTNS HLETEAQSSSL (SEQ ID NO:502). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in Hodgkin's lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, Hodgkin's lymphoma and other immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 285 as residues: Pro-27 to Thr-32.

5 The tissue distribution in Hodgkin's lymphoma and the sequence homology indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of Hodgkin's lymphoma, since the elevated expression and secretion by the tumor mass may be indicative of tumors of this type. Additionally the gene product may be used as a target in the immunotherapy of the cancer. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune
10 functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 53

This gene has extensive homology to cDNA for Homo sapiens mRNA for the
15 ISLR gene(See Accession No. AB003184). This protein is considered to be a new member of the Ig superfamily and contains a leucine-rich repeat (LRR) with conserved flanking sequences and a C2-type immunoglobulin (Ig)-like domain. These domains are important for protein-protein interaction or cell adhesion, and therefore it is possible that the novel protein ISLR may also interact with other proteins or cells. The ISLR gene
20 was mapped on human chromosome 15q23-q24 by fluorescence in situ hybridization (See Medline Article No. 97468140). Homology to the ISLR gene has been confirmed by another independent group as well (See Accession No. Hs.102171)

This gene is expressed in a number of tissues including human retina, heart, skeletal muscle, prostate, ovary, small intestine, thyroid, adrenal cortex, testis,
25 stomach, spinal cord, fetal lung and fetal kidney tissues, colon, tonsil and stomach cancer, and to a lesser extent in endometrial stromal cells treated with estradiol, breast tissue, synovium, lymphoma, and number of other tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
30 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors of colon, ovary and breast origins. However, due to the wide range of expression in various tissues, protein may play a vital role in the development of cancer in other tissues as well, not just those mentioned above. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
35 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the colon, ovary and breast, expression of this gene at significantly higher or lower levels may be routinely

detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Additionally, this gene maps to chromosome 15q23-q24, and therefore, can be used as a marker in linkage analysis for chromosome 15.

The tissue distribution in tumors of colon, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 54

Gene has homology to multidrug resistance gene 1 (See Accession No. P06795). Preferred polynucleotide fragments comprise the following sequence:
 GCTTCGTGTCCAACCCTCTTGCCCTTCGCCTGTGTGCCTGGAGCCAGTCCCA
 CCACGCTCGCGTTTCCTCCTGTAGTGCTCACAGGTCCCAGCACCGATGGCA
 TTCCCTTTGCCCTGAGTCTGCAGCGGGTCCCTTTTGTGCTTCCTTCCCCTCA
 GG TAGCCTCTCTCCCCCTGGGCCACTCCCGGGGGTGAGGGGGTTACCCCTT
 CCCAGTGTTTTTTATTCCTGTGGGGCTCACCCCAAAGTATTAAGTAGCTTT
 GTAA (SEQ ID NO:503). Also preferred are polypeptide fragments encoded by these polynucleotide fragments.

This gene is expressed primarily in lung, esophagus, leukemia (Jurkat cells) and breast cancers and to a lesser extent in macrophages treated with GM-CSF fetal tissues and wide range of tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer of wide range of origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the solid tumors, lung and leukemia, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Furthermore, due to the high expression level in lung tissue and the proposed function of the multidrug resistance protein 1 gene as the efflux pump responsible for low-drug accumulation in multidrug-resistant cells, protein as well mutants thereof, may also be beneficial as a target for gene therapy, particularly for the chronic patient. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 287 as residues: Met-1 to Lys-16.

The tissue distribution in wide range of cancers and fetal tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection of cells in active proliferation, such as cancers. The gene products may be used for cancer markers or immunotherapy target.

FEATURES OF PROTEIN ENCODED BY GENE NO: 55

This gene maps to the X chromosome.

This gene is expressed primarily in the brain and to a lesser extent in the developing embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative disease states and developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders, including sex-linked disorders, of the above tissues or cells, particularly of the neurological, developmental systems, and cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Moreover, this gene maps to the X chromosome, and therefore, may be used as a marker in linkage analysis for this chromosome.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Klinefelter's, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental

disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 56

- 5 The translation product of this gene shares sequence homology with paxillin which is thought to be important in mediating signal transduction from growth factor receptors to the cytoskeleton. Preferred polynucleotide fragments comprise the following sequence: TGGCTCACTGTCTTACAATCACTGCTGTGGAATCATGA TACCACITTTAGCTCTTTGCATCTTCCTTCAGTGTATTTTGTGTTTTCAAGAGG
10 AAGTAGATTTTAACTGGACAACITTTGAGTACTGACATCATTGATAAATAAACT GGCTTGTGGTTTCAA (SEQ ID NO:506). Also preferred are polypeptide fragments encoded by these polynucleotide fragments. More preferably, polypeptide fragments comprise the amino acid sequence: LDELM AHLTEMQAKVAVRAD AGKKHL PDKQDHKASLDSMLGGLEQELQDLGIATVPKGHCASCQKPIAGKVI
15 HALGQSWHPEHFVCTHCKEEIGSSPFFERSGLX YCPNDYHQLFSPRCAYCAAP ILDKVLTAMNQ TWHPEHFFCSHCGEVFGAEGFHEKDKKPYCRKDFLAMFSPK CGGCNRPVLENYLSAMDTVWHPECFVCGDCFTSFSTGSFFELDGRPFCELHYH HRRGTLCHGCGQPITGRCISAMGYKFHPEHFVCAFCLTQLSKGIFREQNDKTY CQPCFNKLF (SEQ ID NO:507); KASLDSMLGGLEQELQDLGIATVPKGHC
20 ASCQKPIAGKVIHAL (SEQ ID NO:508); CPNDYHQLFSPRCAYCAAPILDKVL TAMNQ TWHPEHFFCSHCGEVFGAEG (SEQ ID NO:509); DKKPYCRKDFLAM FSPKCGGCNRPVLENYLSAMDTVWHPECFVCGDCFTSFSTGSFFELDGRPFCE L (SEQ ID NO:510); CGQPITGRCISAMGYKFHPEHFVCAFCLTQLSKGIFRE QNDKTYCQ (SEQ ID NO:511). Polynucleotide fragments encoding these preferred
25 polypeptide fragments are also contemplated.

This gene is expressed primarily in brain, and to a lesser extent in the developing embryo.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
30 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disease states and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and
35 nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or

cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Moreover, since this gene shares homology with a gene that maps to chromosome 11, (See Accession No.T87404), gene as well as its translated product may be used for linkage analysis on chromosome 11.

The tissue distribution and homology to paxillin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and or detection of disease states associated with abnormal signal transduction in brain and/or the developing embryo. This would include treatment or detection of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder and also in the treatment and or detection of embryonic development defects.

FEATURES OF PROTEIN ENCODED BY GENE NO: 57

This gene is expressed primarily in fetal spleen, brain, and to a lesser extent in six week old embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders, neurological disorders, and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and developmental systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 290 as residues: Arg-28 to Gly-34.

The expression of this gene in fetal spleen indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/detection of immune disorders such as arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. In addition the expression of this gene in the early embryo, indicates a key role in embryo development and hence the gene or gene product could be used in the treatment and or detection of embryonic development defects. This would include

treatment or detection of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder and also in the treatment and or detection of embryonic development defects.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 58

The translation product of this gene shares sequence homology with the gene disrupted in the neurodegenerative disease dentatorubal-pallidolusian atrophy. Moreover a long open reading frame exists in an alternative frame. Preferred polypeptide fragments

10 comprise the following:

MGSSQSVEIPGGGTEGYHVLRVQENSPGHRAGLEPFFDFIVSINGSRLNKDND
 TLKDLLKXNVEKPVKMLIYSSKTLELRETSVTPSNLWGGQGLLGVSIRFCSFD
 GANENVWHVLEVESNSPAALAGLRPHSDYIIGADTMNESEDLSLIETHEAKP
 LKLYVYNTDNDNCREVIITPNSAWGGEGSLGCGIGYGYLHRIPTRPFEEGKKIS
 15 LPGQMAGTPITPLKDGFTVEVQLSSVNPPSLSPPGTTGIEQSLTGLSISSTPPAVSS
 VLSTGVPTVPLLPPQVNQSLTSVPPMNPATTLPGLMPLPAGLPNLPNLPNLP
 PHIMPGVGLPELVNPGLPPLPSMPPRNLPGIAPLPLPSEFLPSFPLVPESSSAASS
 GELLSSLPPTSNA PSDPATTTAKADAASSLTVDVTPPTAKAPTTVEDRVGDSTPV
 SEKPVSAAVDANASESP (SEQ ID NO:512); SVEIPGGGTEGYHVLRVQENSPGH
 20 RAGLEPFFDFIVSINGSRLNKDNDTLKDLLKXNVEKPVKMLIYSSKTLELRETS
 VTPSNLWGGQGLLGVSIRFCSFDGANENVWH (SEQ ID NO:513); ESNPAA
 LAGLRPHSDYIIGADTMNESEDLSLIETHEAKPLKLYVYNTDNDNCREVIITP
 NSAWGGEGSLGCGIGYGYLHRIPTRPFEEGKKISLPGQMAGTPITPLKDGFTVEV
 QLSSVNPPSLSPPGTTGIEQSLTGLSISSTPPAVS (SEQ ID NO:514); RIPTRPFEEGKKI
 25 SLPGQMAGTPITPLKDGFTVEVQLSSVNPPSLSPPGTTGIEQSLTGLSISSTPPAVS
 SVLSTGVPTVPLLPPQVNQSLTSVPPMNPATTLPGLMPLPAGLPNLPNLPNLP
 APHIMPGVGLPELVNPGLPPLPSMPPRN (SEQ ID NO:516); PGLPPLPSMPPRN
 LPGIAPLPLPSEFLPSFPLVPESSSAASSGELLSSLPPTSNA PSDPATTTAKADAA
 SSLTVDVTPPTAKAPTTVEDRVGDSTPVSEKPVSAAVDAN (SEQ ID NO:517).

30 This gene is expressed primarily in prostate cancer, and to a lesser extent in the pineal glands and in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological conditions and pulmonary disorders. Similarly,
 35 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For

a number of disorders of the above tissues or cells, particularly of the nervous, pulmonary, and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 291 as residues: Asn-9 to Leu-14.

The abundance of this gene in the pineal gland and its homology to a gene disrupted in the neurodegenerative disease state Dentatorubral-pallidoluysian atrophy indicates that this gene may be useful in the treatment and/or detection of other neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. The abundance of this gene in fetal lung would suggest that misregulation of the expression of this protein product in the adult could lead to lymphoma or sarcoma formation, particularly in the lung; that it may also be involved in predisposition to certain pulmonary defects such as pulmonary edema and embolism, bronchitis and cystic fibrosis; and thus the gene or the gene protein encoded by the gene could be used in the detection and/or treatment of these pulmonary disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 59

This gene is expressed primarily in the developing embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developmental system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The expression of this gene primarily in the embryo, indicates the gene plays a key role in embryo development and that the gene or the protein encoded by the gene could be used in the treatment and or detection of developmental defects in the embryo or in infants.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 60

This gene displays homology to nestin, an intermediate filament protein, the expression of which correlates with the proliferation of Central Nervous System progenitor cells and that is useful in the identification of brain tumors. This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1 (See Accession No. AA527348).

This gene is expressed primarily in kidney and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, renal disorders and neurodegenerative conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the excretory and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 293 as residues: Thr-128 to Asn-135.

The tissue distribution and homology to nestin indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and/or treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, its abundance in kidney indicates that it is useful in the treatment and detection of acute renal failure and other disease states associated with the kidney.

FEATURES OF PROTEIN ENCODED BY GENE NO: 61

Gene shares homology with the latrophilin-related protein 1 precursor as well as the calcium-independent alpha-latrotoxin receptor. Preferred polypeptide fragments

comprise the following amino acid sequence:

IYKVFRHTAGLKPEVSCFENIRSCARXXXXXXXXXXXXXWIFGVLHVHVASVV
TAYLFTVSNAFQGMFIFLFLCVLSRKIQEEYYRLFKNVPCC (SEQ ID NO:518);
WIFGVLHVHVASVVTA YLFTVSNAFQGMFIFLFLCVLSRKIQEEYYRLFKNVPC

- 5 C (SEQ ID NO:519). Also preferred are polynucleotide fragments encoding these polypeptide fragments. (See Accession No. 2213659) The translation product of this gene shares sequence homology with CD 97, a seven transmembrane bound receptor.

This gene is expressed primarily in infant brain and in endothelial cells.

- Therefore, polynucleotides and polypeptides of the invention are useful as
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For
15 a number of disorders of the above tissues or cells, particularly of the neurological and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the
20 standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 294 as residues: Lys-13 to Leu-21.

- The tissue distribution of this gene suggest that it may be useful in the detection and/or treatment of neurodegenerative disease states and behavioral disorders such as
25 Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder, while its expression in hematopoietic cell types indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, asthma and immunodeficiency diseases.

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 62

This gene is expressed primarily in fetal liver and fetal spleen.

- Therefore, polynucleotides and polypeptides of the invention are useful as
35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematological and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 295 as residues: Ser-91 to Lys-98.

The tissue distribution of this gene fetal liver and spleen indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, leukemia, asthma and immunodeficiency diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 63

Gene shares homology with human serum amyloid protein. Preferred polypeptide fragments comprise the following amino acid sequence:
ALTRIPPGDWVINVTAVSFAGKTTARFFHSSPPSLGDQARTDPGHQRRD (SEQ ID NO:520) (See Accession No. W13671). Also preferred are polynucleotide fragments encoding these polypeptide fragments This gene maps to chromosome 9, and therefore, may be used as a marker in linkage analysis for chromosome 9 (See Accession No. AA004342).

This gene is expressed primarily in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene in fetal liver-spleen indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, leukemia, asthma, and immunodeficiency diseases.

5 **FEATURES OF PROTEIN ENCODED BY GENE NO: 64**

This gene maps to chromosome 3, and therefore, may be used as a marker in linkage analysis for chromosome 3 (See Accession No. AA219669).

This gene is expressed specifically in the brain.

Therefore, polynucleotides and polypeptides of the invention are useful as
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative disease states. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of
15 the above tissues or cells, particularly of the neurological systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the
20 expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's
25 Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 65

Gene shares homology with a yeast protein. Preferred polypeptide fragments
30 comprise the following amino acid sequence: LQEVNITLPENSVWYERYKFDIP VFHL (SEQ ID NO:521). Also preferred are polynucleotide fragments encoding these polypeptide fragments. (See Accession No. 1332638)

This gene is expressed primarily in fetal tissue (fetus and fetal liver).

Therefore, polynucleotides and polypeptides of the invention are useful as
35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, liver disorders and cancers (e.g. hepatoblastoma). Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 298 as residues: Asn-59 to Glu-64.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 66

Gene has homology with a B-cell surface antigen which may indicate gene plays a role in the immune response, including, but not limited to disorders and infections of the immune system. Preferred polynucleotide fragments comprise the following sequence: TAGCATGTAGCCAGTCGAATAACNTATAAGGACAAAGTGGAGTC CACGCGTGCGGCCGTCTAGACTAGTGGATCCCCGGCTGCAGGATTCGGC ACGAG (SEQ ID NO:523). Also preferred are polypeptide fragments encoded by these polynucleotide fragments (See Accession No.T94535). Additionally, this gene shares homology with an interferon-gamma receptor. Preferred polypeptide fragments also comprise the following amino acid sequence: MQGSGSQFRACLLCLCFSCPC SPGGPRWNSRQGGRFPKTCRAISQNLVFKYKTFCPVRYMQPHRSSLCLHFTS YVFILSTWGSRLTYSTDLLKKKKNSRGGPVPIRPKS (SEQ ID NO:522); MQGSGSQFRACLLCLCFSCPCSPGGPRWNSRQGGRFPKTCRAISQNLVFK (SEQ ID NO:524); PVRYMQPHRSSLCLHFTSYVFILSTWGSRLTYSTDLLKKKK NSRGGPVPIRPKS (SEQ ID NO:525); and GEEQRDCSLGWRGVGMRATHCQAA RMFVLFSLPKYAGL (SEQ ID NO:526). Also preferred are polynucleotide fragments encoding these polypeptide fragments

This gene is expressed primarily in T-cells and gall bladder.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological disorders and conditions (immunodeficiencies, cancer, leukemia, hematopoiesis). Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and digestive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 299 as residues: Thr-41 to Gly-52.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of immune disorders including: leukemias, lymphomas, auto-immune disorders, immunosuppressive (transplantation) and immunodeficiencies (e.g. AIDS), inflammation and hematopoietic disorders. The expression of this gene in gall bladder would suggest a possible role for this gene product in digestive disorders, particularly of the pancreas.

FEATURES OF PROTEIN ENCODED BY GENE NO: 67

This gene maps to chromosome 11, and therefore, may be used as a marker in linkage analysis for chromosome 11 (See Accession No. AA011622).

This gene is expressed primarily in a variety of fetal and developmental tissues (e.g. fetal spleen, infant brain).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental, immune or neurological abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing immune and central nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 300 as residues: Ser-38 to Ser-43.

5 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for developmental abnormalities or fetal deficiencies. The detection in infant brain would suggest a role in neurological disorders (both developmental and neurodegenerative conditions of the brain and nervous system, behavioral disorders, depression, schizophrenia, Alzheimer's disease, Parkinson's
10 disease, Huntington's disease, mania, dementia). In addition, the detection in spleen would similarly suggest a role in detection and treatment of immunologically mediated disorders (e.g. immunodeficiency, inflammation, cancer, wound healing, tissue repair, hematopoiesis).

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 68**

 This gene is expressed primarily in spleen, T-cells, and fetal heart.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
20 not limited to, immunological deficiencies, including AIDS and cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and cardiovascular systems, expression of this gene at significantly higher
25 or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, autoimmune disorders, immunodeficiencies (e.g. AIDS), immuno-suppressive conditions (transplantation) and hematopoietic disorders. The expression in fetal heart indicates that polynucleotides and
35 polypeptides corresponding to this gene are useful for the treatment and diagnosis of cardiovascular disorders (e.g. heart disease, restenosis, atherosclerosis, stroke, angina, thrombosis).

FEATURES OF PROTEIN ENCODED BY GENE NO: 69

Gene shares homology with a human collagen protein. Preferred polypeptide fragments comprise the following amino acid sequence:

- 5 MPRKTSKCRQLLCGASRNADTAARQSTCSSHRPPGKIPSLGPRRXPGCXSV
SSRGEQSTGSPAAPRCGRRDAHRGLPGGAAMTPGDTWASFNPRAGHSKSQGE
GQESSGASRQDRHPVSHWVERQREAWGAPRSSSAGGVKVAATTEREPEFKIK
TGKA (SEQ ID NO:527); CSGASRNADTAARQSTCSSHRPPGKIPSLGPRRXPG
CXSVSSRGEQSTGSPAAPRCGRRDAHRGLPGGAAMTPGDTWASFNPRAGHS
10 (SEQ ID NO:528); QGEGQESSGASRQDRHPVSHWVERQREAWGAPRSSSAGG
VKVAATTEREPEFKIKTGKA (SEQ ID NO:529) (See Accession No. 124886). Also
preferred are polynucleotide fragments encoding these polypeptide fragments

This gene is expressed primarily in fetal heart.

- Therefore, polynucleotides and polypeptides of the invention are useful as
15 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, cardiovascular disorders. Similarly, polypeptides and antibodies directed
to these polypeptides are useful in providing immunological probes for differential
identification of the tissue(s) or cell type(s). For a number of disorders of the above
20 tissues or cells, particularly of the cardiovascular system, expression of this gene at
significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
fluid or spinal fluid) or another tissue or cell sample taken from an individual having
such a disorder, relative to the standard gene expression level, i.e., the expression level
25 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred
epitopes include those comprising a sequence shown in SEQ ID NO: 302 as residues:
Pro-32 to Ser-39.

- The tissue distribution indicates that polynucleotides and polypeptides
corresponding to this gene are useful for the treatment and diagnosis of cardiovascular
30 disorders (e.g. heart disease, restenosis, atherosclerosis, stroke, angina, thrombosis).

FEATURES OF PROTEIN ENCODED BY GENE NO: 70

The translation product of this gene shares sequence homology with a chicken
single-strand DNA-binding protein. Preferred polypeptide fragments comprise the
35 following amino acid sequence:

MSPRYPGGPRPPLRIPNQALGGVPGSQPLLPSGMDPTRQQGHPNMGGPMQRM
TPPRGMVPLGPQNYGGAMRPPLNALGGPGMPGMNMGPGGGRPWPNTNAN

SIPYSSASPGNYVGPPGGGGPPGTPIMPSPADSTNSGDNMYTLMNAVPPGPNR
 PNFPMPGSDGPMGGLGGMESHMNGSLGSGDMDSISKNSPNNMSLSNQ
 GTPRDDGEMGGNFLNPFQSESYSPSMTMSV (SEQ ID NO:530); MSPRYPGG
 PRPPLRIPNQALGGVPGSQPLLPSGMDPTRQQGHPNMGGPMQRMTPPRGMVP
 5 LGPQNYGGAMRPPLNALGGPGMPGMNMGPGGGRPWPNTNANSIPYSSASP
 GNY (SEQ ID. NO:531); LNALGGPGMPGMNMGPGGGRPWPNTNANSIPYSS
 ASPGNYVGPPGGGGPPGTPIMPSPADSTNSGDNMYTLMNAVPPGPN (SEQ ID
 NO:532); GPMGGLGGMESHMNGSLGSGDMDSISKNSPNNMSLSNQPGTPR
 DDGEMGGNFLNPFQSESYSPSMTMSV (SEQ ID NO:533); TCEHSSEAKAFHDY
 10 (SEQ ID NO:534). Also preferred are polynucleotide fragments encoding these
 polypeptide fragments. (See Accession No. 1562534)

This gene is expressed primarily in placenta and to a lesser extent in the fetal heart and a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as
 15 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, developmental abnormalities, fetal deficiencies, and particularly of the
 cardiovascular system. Similarly, polypeptides and antibodies directed to these
 polypeptides are useful in providing immunological probes for differential identification
 20 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,
 particularly of the reproductive system, expression of this gene at significantly higher or
 lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
 another tissue or cell sample taken from an individual having such a disorder, relative to
 25 the standard gene expression level, i.e., the expression level in healthy tissue or bodily
 fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
 corresponding to this gene are useful for the detection and treatment of developmental
 abnormalities or fetal deficiencies, ovarian and other endometrial cancers, reproductive
 30 dysfunction, cardiovascular disorders, and pre-natal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 71

This gene is expressed primarily in fetal liver and to a lesser extent in the breast and testes.

35 Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, liver disorders (including hepatoblastomas) and reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). The expression in testes and breast indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of endocrine and reproductive disorders (e.g. sperm maturation, milk production, testicular and breast cancers).

FEATURES OF PROTEIN ENCODED BY GENE NO: 72

This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1 (See Accession No. W93595).

This gene is expressed primarily in smooth muscle and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiovascular and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular and central nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of restenosis, atherosclerosis, stroke, angina, thrombosis, wound healing and other conditions of heart disease. In addition, the expression in brain would suggest that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of developmental, degenerative and behavioral conditions of the brain and nervous system (e.g. schizophrenia, depression, Alzheimer's disease, Parkinson's disease, Huntington's disease, mania, dementia, paranoia, addictive behavior and sleep disorders).

FEATURES OF PROTEIN ENCODED BY GENE NO: 73

Gene shares homology with human stromalin-2. Preferred polypeptide fragments comprise the following amino acid sequence:

QAFVLLSLLLLIFSPQMIVGGRDFLRPLVFFPEATLQSELASFLMDHVFQPGDL
GSGA (SEQ ID NO:535); ACSYLLCNPEFTFFSRADFARSQLVDLLTDRFQQE
LEELLQVG (SEQ ID NO:536); QKQLSSLRDRMVAFCCLCQSCSDVDTEIQEQV
ST (SEQ ID NO:537); QVILPALTLVYFSILWTLTHISKSDAS (SEQ ID NO:538);
STHDLTRWELYEPCCQLLQKAVDTGXVPHQV (SEQ ID NO:539). Also preferred
are polynucleotide fragments encoding these polypeptide fragments (See Accession
No.R65208) This gene maps to chromosome 7, and therefore, may be used as a
marker in linkage analysis for chromosome 7 (See Accession No. D52585).

This gene is expressed primarily in the brain (infant brain, adult brain, pituitary, cerebellum, hippocampus, schizophrenic hypothalamus, amygdala).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental and neurodegenerative diseases of the brain and nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those

comprising a sequence shown in SEQ ID NO: 306 as residues: Thr-25 to Lys-36, Lys-55 to Ser-63.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of developmental, degenerative and behavioral conditions of the brain and nervous system (e.g. schizophrenia, depression, Alzheimer's disease, Parkinson's disease, Huntington's disease, mania, dementia, paranoia, addictive behavior and sleep disorders).

FEATURES OF PROTEIN ENCODED BY GENE NO: 74

This gene is expressed primarily in the hypothalamus of a human suffering from schizophrenia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the CNS particularly schizophrenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS, such as schizophrenia expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 307 as residues: Gly-38 to Ala-44.

The tissue distribution indicates that the protein products of this gene are useful for the study, diagnosis and treatment of schizophrenia and other disorders involving the CNS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 75

Preferred polypeptides of the invention comprise the following amino acid sequence encoded by this gene:

LAVSTSFICCADISTALPLGSSRPAPAPRHREHEHGHQARPPRLXLSLMPLSTP
AAAQLLWTQLTPMGGRPGGRHSPTLHTGPRALPPGPPHPSLHVAALSLLR

(SEQ ID NO:540). Polynucleotides encoding such polypeptides are also provided.

This gene is expressed primarily in endometrial tumor and to a lesser extent in amniotic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, reproductive and immune disorders particularly cancers of those systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 308 as residues: Ser-3 to Arg-9.

The tissue distribution indicates that the protein products of this gene are useful for study and treatment of immune and reproductive disorders particularly cancers of those systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 76

This gene is expressed primarily in kidney cortex and to a lesser extent in early stage human brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, renal disorders such as renal cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the kidney expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 309 as residues: Gly-38 to Gly-45, Gly-47 to Gly-52, Pro-92 to Lys-110.

The tissue distribution indicates that the protein products of this gene are useful for study, treatment and diagnosis of renal diseases such as cancer of the kidney.

FEATURES OF PROTEIN ENCODED BY GENE NO: 77

This gene is expressed primarily in kidney medulla.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, metabolic and renal disorders. Similarly, polypeptides and antibodies
directed to these polypeptides are useful in providing immunological probes for
10 differential identification of the tissue(s) or cell type(s). For a number of disorders of
the above tissues or cells, particularly of the metabolic and renal systems, expression of
this gene at significantly higher or lower levels may be routinely detected in certain
tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
15 individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

The tissue distribution indicates that the protein products of this gene are useful
for study, treatment and diagnosis of metabolic and renal diseases and disorders.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 78

This gene is expressed in chronic synovitis and microvascular endothelium.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
25 not limited to, arthritis and atherosclerosis. Similarly, polypeptides and antibodies
directed to these polypeptides are useful in providing immunological probes for
differential identification of the tissue(s) or cell type(s). For a number of disorders of
the above tissues or cells, particularly of the vascular and skeletal systems, expression
of this gene at significantly higher or lower levels may be routinely detected in certain
30 tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

35 The tissue distribution indicates that the protein products of this gene are useful
for study, diagnosis and treatment of arthritic and other inflammatory diseases as well
as cardiovascular diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 79

This gene is expressed in resting T-cells and activated monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, immune disorders. Similarly, polypeptides and antibodies directed to
these polypeptides are useful in providing immunological probes for differential
10 identification of the tissue(s) or cell type(s). For a number of disorders of the above
tissues or cells, particularly of the immune system, expression of this gene at
significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
fluid or spinal fluid) or another tissue or cell sample taken from an individual having
15 such a disorder, relative to the standard gene expression level, i.e., the expression level
in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful
for the study and treatment of immune diseases such as inflammatory conditions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 80

20 This gene is expressed in a variety of immune system tissues, e.g., neutrophils,
T-cells, and TNF induced epithelial and endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
25 not limited to, infectious and immune disorders. Similarly, polypeptides and antibodies
directed to these polypeptides are useful in providing immunological probes for
differential identification of the tissue(s) or cell type(s). For a number of disorders of
the above tissues or cells, particularly of the immune and vascular systems, expression
of this gene at significantly higher or lower levels may be routinely detected in certain
30 tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
35 NO: 313 as residues: Met-1 to Trp-6.

The tissue distribution indicates that the protein products of this gene are useful
for study and treatment of infectious diseases, immune and vascular disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 81

This gene is expressed in activated neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, inflammation and other immune conditions. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
10 for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the immune system, expression of this gene
at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
fluid or spinal fluid) or another tissue or cell sample taken from an individual having
15 such a disorder, relative to the standard gene expression level, i.e., the expression level
in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful
for study and treatment of immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 82

20 This gene is expressed in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, inflammatory and other immune conditions. Similarly, polypeptides and
25 antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the immune system, expression of this gene
at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
30 fluid or spinal fluid) or another tissue or cell sample taken from an individual having
such a disorder, relative to the standard gene expression level, i.e., the expression level
in healthy tissue or bodily fluid from an individual not having the disorder. Preferred
epitopes include those comprising a sequence shown in SEQ ID NO: 315 as residues:
Ala-83 to Thr-91.

35 The tissue distribution indicates that the protein products of this gene are useful
for study and treatment of immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 83

This gene is expressed in human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, inflammation and immune disorders. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
10 for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the immune and inflammatory system,
expression of this gene at significantly higher or lower levels may be routinely detected
in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
15 an individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

The tissue distribution indicates that the protein products of this gene are useful
for diagnosis and treatment of disorders of the inflammatory and immune systems.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 84

This gene is expressed in human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
25 not limited to, disorders of the inflammatory and immune systems. Similarly,
polypeptides and antibodies directed to these polypeptides are useful in providing
immunological probes for differential identification of the tissue(s) or cell type(s). For
a number of disorders of the above tissues or cells, particularly of the inflammatory and
immune systems, expression of this gene at significantly higher or lower levels may be
30 routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily
fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or
cell sample taken from an individual having such a disorder, relative to the standard
gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
individual not having the disorder.

35 The tissue distribution indicates that the protein products of this gene are useful
for diagnosis and treatment of disorders of the immune and inflammatory systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 85

This gene is expressed in activated neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, inflammation and immune system diseases. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
10 of the above tissues or cells, particularly of the immune system and inflammatory
system, expression of this gene at significantly higher or lower levels may be routinely
detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g.,
serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
taken from an individual having such a disorder, relative to the standard gene
15 expression level, i.e., the expression level in healthy tissue or bodily fluid from an
individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful
for diagnosis and treatment of diseases of the inflammatory and immune systems.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 86

This gene is expressed in activated neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
25 not limited to, inflammation and immune system disorders. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the inflammatory and immune system,
expression of this gene at significantly higher or lower levels may be routinely detected
30 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
an individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
35 NO: 319 as residues: Met-1 to Gly-6, Gly-32 to Pro-43, Leu-55 to Gln-60.

The tissue distribution indicates that the protein products of this gene are useful
for diagnosis and treatment of disorders of the immune and inflammatory system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 87

In specific embodiments, polypeptides of the invention comprise the sequence:

EQVLALLWPRFELILEMNVQSVRSTDPQRLGGLDTRPHYTTRRYAEFSSALVSIN
 5 QTIPNERTMQLLGQLQVEVENFVLRVAAEFSSRKEQLVFLINNYDMMLGVLMER
 RAADDKEVESFQQLLNARTQEFIEELLSPFPGGLVAFVKEAEALIERGQAERLR
 GEEARVTQLIRGFGSSWKSSVESLSQDVMRSFTNFRNGTSIIQG (SEQ ID
 NO:541),ALLKYRFFYQFLGNERATAKEIRDEYVETLSKIYLSYYRSYLGRLMK
 VQYEEVAEKDDLGMGVEDTAKKGFXSKPSRSRNTIFTLGTRGSVISPTELEAPILV
 10 PHTAQR (SEQ ID NO: 542); EQRYPFALFRSQHYXLLDNSCREYLFICEFFVVS
 GPXAHDLFHAVMGRTLSMTLKHLDSYLADCYDAIAVFLCIHIVLRFRNIAAKRD
 VPALDRYW (SEQ ID NO:543),GGLDTRPHYTTRRYAEFSSALVSINQ (SEQ ID
 NO:544); SRKEQLVFLINNYDMMLGV (SEQ ID NO: 545) and/or ALLKYRFFY
 QFLGNERATAKEIRDEYVETLSKIYLSYYRSYLGRLMKVQYEEVAEKDDLGMG
 15 VEDTAKKGFXSKPSLRSRNTIFTLGTRGSVISPTELEAPILVPHTAQRXEQRYPF
 EALFRSQHYXLLDNSCREYLFICEFFVVS GPXAHDLFHAVMGRTLSMTLKHL
 SYLADCYDAIAVFLCIHIVLRFRNIAAKRDVPALDRYWEQVLALLWPRFELILEM
 NVQSVRSTDPQRLGGLDTRPHYTTRRYAEFSSALVSINQTIPNERTMQLLGQLQV
 EVENFVLRVAAEFSSRKEQLVFLINNYDMMLGVLMERAAADDKEVESFQQLLN
 20 ARTQEFIEELLSPFPGGLVAFVKEAEALIERGQAERLRGEEARVTQLIRGFGSSW
 KSSVESLSQDVMRSFTNFRNGTS (SEQ ID NO:546). Polynucleotides encoding
 these polypeptides are also encompassed by the invention. The translation product of
 this gene shares sequence homology with suppressor of actin mutation which is thought
 to be important in mutation suppression.

25 This gene is expressed primarily in fetal liver and to a lesser extent in a variety
 of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 30 not limited to, liver and mutations. Similarly, polypeptides and antibodies directed to
 these polypeptides are useful in providing immunological probes for differential
 identification of the tissue(s) or cell type(s). For a number of disorders of the above
 tissues or cells, particularly of the liver or cancer, expression of this gene at
 significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
 35 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
 fluid or spinal fluid) or another tissue or cell sample taken from an individual having
 such a disorder, relative to the standard gene expression level, i.e., the expression level

in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 320 as residues: Val-53 to Arg-60, Thr-88 to Thr-94, Ala-142 to Ser-150, Gly-188 to Glu-196, Gly-208 to Ser-214, Thr-227 to Gly-232, Lys-279 to Phe-285.

- 5 The tissue distribution and homology to suppressor of actin mutation suggest that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and of liver disorder or cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 88

- 10 This gene maps to chromosome 9, and therefore can be used in linkage analysis as a marker for chromosome 9. In specific embodiments, polypeptides of the invention comprise the sequence:

- YEGKEFDYVFSIDVNEGGPSYKLPYNTSDDPWLTAYNFLQKNDLNPMFLDQVA
KFIIDNTKGQMLGLGNPSFSDPFTGGGRYVPGSSGSSNTLPTADPFTGAGRYV
15 PGASMGTTMAGVDPFTGNSAYRSAASKTMNIYFPKKEAVTFDQANPTQILGK
LKELNGTAPEEKKLTEDDLILLEKILSLICNSSSEKPTVQQLQILWKAINCPEDIV
FPALDILRLSIKHPSVNENFCNEKEGAQFSSHLLNLLNPKGKPANQLLALRTFC
NCFVGQAGQKLMMSQRESLSHAIELKSGSNKNI (SEQ ID NO: 547);
HIALATLALNYSVCFHKD (SEQ ID NO: 548); HNIEGKAQCLSLISTILEVVQ
20 DLEATFRLLVALGTLISDDSNVQLAKS (SEQ ID NO: 549); LGVDSQIKKYSS
VSEPAKVSECCRFILNLL (SEQ ID NO: 550); and/or YEGKEFDYVFSIDVNEGGPS
YKLPYNTSDDPWLTAYNFLQKNDLNPMFLDQVAKFIIDNTKGQMLGLGNPSFS
DPFTGGGRYVPGSSGSSNTLPTADPFTGAGRYVPGASMGTTMAGVDPFTGN
SAYRSAASKTMNIYFPKKEAVTFDQANPTQILGKLKELNGTAPEEKKLTEDDLI
25 LLEKILSLICNSSSEKPTVQQLQILWKAINCPEDIVFPALDILRLSIKHPSVNENFC
NEKEGAQFSSHLLNLLNPKGKPANQLLALRTFCNCFVGQAGQKLMMSQRESL
MSHAIELKSGSNKNIHIALATLALNYSVCFHKDHNIEGKAQCLSLISTILEVVQD
LEATFRLLVALGTLISDDSNVQLAKSLGVDSQIKKYSSVSEPAKVSECCRFILN
LL (SEQ ID NO: 551). Polynucleotides encoding these polypeptides are also
30 encompassed by the invention. These polypeptides share significant homology with
phospholipase A2 activating protein which is thought to be important in signal
transduction (see, e.g., Wang et al., Gene 161(2):237-241 (1995)).

- This gene is expressed primarily in endothelial cells, to a less extent in placenta,
endometrial stromal cells, osteosarcoma, testis tumor, muscle, and infant brain that are
35 likely to be rich in blood vessels.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders in vascular system, aberrant angiogenesis, tumor angiogenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system or tumors, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene in endothelial cells and several potential highly vascularized tissues and its homology to phospholipase A2 activating protein suggest that this gene may be involved in transducing signals for endothelial cells in angiogenesis or vasculogenesis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 89

In specific embodiments, polypeptides of the invention comprise the sequence: YPNQDGDILRDQVLHEHIQRLSKVVTANHRALQIPEVYLREAPWPSAQSEIRTIS
 20 AYKTPRDKVQCILRMCSTIMNLLSLANEDSVPGADDFVPVLVFLIKANPPCLL
 STVQYISSFYASCLSGEESYWWMQFTA AVE (SEQ ID NO:552); YPNQDGDILR
 DQVLHEHIQRLSKVVTANHRALQIPEVYLREAPWPSAQSEIRTISAYKTPRDKVQ
 CILRMCSTIMNLLSLANEDSVPGADDFVPVLVFLIKANPPCLLSTVQYISSFYA
 SCLSGEESYWWMQFTA AVEFIKTI (SEQ ID NO:553); YPNQDGDILRDQVL (SEQ
 25 ID NO:554); EAPWPSAQSEI (SEQ ID NO:555); PVLVFLIKANP (SEQ ID
 NO:560); SGEESYWWMQFTA AVEFIKTI (SEQ ID NO:556); ADDFVPVLVF
 VLIKANPP (SEQ ID NO:557); YKTPRDKVQCIL (SEQ ID NO:558); and/or
 GADDFVPVLVFLIK (SEQ ID NO:559). The translation product of this gene shares
 sequence homology with human ras inhibitor and yeast VPS9p which is thought to be
 30 important in golgi vacuole transport.

This gene is expressed primarily in T cells and melanocytes and to a lesser extent in a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, dysfunction and disorders involving T cells and melanocytes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ras inhibitor indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating signal transduction; diagnosis and treatment of disorders involving T cells and melanocytes.

FEATURES OF PROTEIN ENCODED BY GENE NO: 90

This gene maps to chromosome 9 and therefore polypeptides of the invention can be used in linkage analysis as a marker for chromosome 9. The translation product of this gene shares sequence homology with neuronal olfactomedin-related ER localized protein which is thought to be important in influence the maintenance, growth, or differentiation of chemosensory cilia on the apical dendrites of olfactory neurons. In specific embodiments, polypeptides of the invention comprise the sequence: SARASTQPPAGQHGPC (SEQ ID NO:561); MPGRWRWQRDMHPARKLLSL FLILMGTELTQD (SEQ ID NO:562); SAAPDSLLRSSKGSTRGSL (SEQ ID NO:563); AAIVTWRGKSESRIAKTPGI (SEQ ID NO:564); FRGGGTLVLPPTHT PEWLIL (SEQ ID NO:567); PLGITLPLGAPETGGGD (SEQ ID NO:565); and/or CAAETWKGSQRAGQLCALLA (SEQ ID NO:566).

This gene is expressed in pineal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological and endocrinological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neurological or endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 323 as residues: Leu-20 to Ala-26, Arg-32 to Arg-39, Thr-104 to Gly-112.

- 5 The tissue distribution and homology to olfactomedin-related protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for maintenance, growth, or differentiation of neuron cells in pineal gland, therefore, may be useful for diagnosis and treatment of neurological disorders in pineal gland.

FEATURES OF PROTEIN ENCODED BY GENE NO: 91

- 10 This gene is expressed primarily in prostate and apoptotic T cells.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, prostate disease and T cell dysfunction. Similarly, polypeptides and
15 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate cancer, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
20 fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detect abnormal activity in prostate and T cells
25 or probably treatment of this abnormality.

FEATURES OF PROTEIN ENCODED BY GENE NO: 92

This gene is expressed primarily in prostate and to a lesser extent in smooth muscle cells, fibroblasts, and placenta.

- 30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders in prostate or vascular system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes
35 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prosate or vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain

tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating function of prostate or highly vascularized tissues, e.g. placenta.

10 **FEATURES OF PROTEIN ENCODED BY GENE NO: 93**

This gene is expressed primarily in embryos and fetal tissues stage human and to a lesser extent in a wide variety of other proliferative tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders in embryonic development and cell proliferation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryonic tissues and proliferative cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of abnormalities in developing and proliferative cells and organs.

30 **FEATURES OF PROTEIN ENCODED BY GENE NO: 94**

The translation product of this gene shares sequence homology with transformation related protein which is thought to be important in transformation.

This gene is expressed primarily in female reproductive tissues, i.e., breast cancer cells, placenta, and ovary and to a lesser extent in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, cancer or dysfunction of reproductive tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproduction system,

- 5 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the
10 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 327 as residues: Ser-50 to Pro-61.

- The tissue distribution and homology to transformation related protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of conditions caused by transformation, i.e. tumorigenesis in
15 reproductive organs, e.g. breast, placenta, and ovary.

FEATURES OF PROTEIN ENCODED BY GENE NO: 95

This gene is expressed primarily in testes, rhabdomyosarcoma, infant brain and to a lesser extent in some tumors and highly vascularized tissues.

- 20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumorigenesis, abnormal angiogenesis, and/or neurological disorders. , Similarly, polypeptides and antibodies directed to these polypeptides are useful in
25 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumor tissues or vascular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
30 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 328 as residues: Arg-46 to Trp-54, Pro-60 to Ile-69, Asn-116 to Ala-122, Arg-147 to Lys-153, Ser-158 to Glu-170, Ile-399 to
35 Ser-405, Pro-486 to Met-499, Pro-502 to Asp-508.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for a range of disease states including treatment of

tumor or vascular disorders and the treatment of neurological disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 96

This gene maps to chromosome 7 and therefore polynucleotides of the present invention can be used in linkage analysis as a marker for chromosome 7. The translation product of this gene is homologous to the *Clostridium perfringens* enterotoxin (CPE) receptor gene product and shares sequence homology with a human ORF specific to prostate and a glycoprotein specific to oligodendrocytes both of which are tissue specific proteins. (See e.g., Katahira et al., J Cell Biol. 136(6):1239-1247 (1997). PMID: 9087440; UI: 97242441.

This gene is expressed primarily in pancreas tumor and ulcerative colitis and to a lesser extent in several tumors and normal tissues.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pancreatic disorder, ulcerative colitis, tumors and food poisoning. Similarly, polypeptides and antibodies directed to these polypeptides are useful in
20 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system or tumorigenic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal
25 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 329 as residues: Gly-147 to Met-152, Cys-177 to Lys-188.

30 The tissue distribution and homology to prostate and oligodendrocyte-specific protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for marker of diagnosis or treatment of disorder in pancreas, ulcerative colitis, and tumors. Furthermore, identity to the human receptor for *Clostridium perfringens* enterotoxin indicates that the soluble portion of this receptor could be used in the
35 treatment of food poisoning associated with *Clostridia perfringens* by blocking the activity of *perfringens* enterotoxin.

FEATURES OF PROTEIN ENCODED BY GENE NO: 97

The translation product of this gene shares sequence homology with ATPase which is thought to be important in metabolism.

5 This gene is expressed primarily in testes and several hematopoietic cells and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, leukemia and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 330 as residues: Leu-37 to Ala-42.

The tissue distribution and homology to ATPase indicates that polynucleotides and polypeptides corresponding to this gene are useful for marker of diagnosis and treatment of leukemia and other hematopoietic disorders.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 98

In specific embodiments, polypeptides of the invention comprise the sequence: MRSARPSLGCLPSWAFSQALNI (SEQ ID NO:568); LLGLKGLAPAEISAVCE KGNFN (SEQ ID NO:569); VAHGLAWSYYIGYLRILPELQARIR (SEQ ID NO:570); TYNQHYNLLRGAVSQRC (SEQ ID NO:571); ILLPLDCGVPDNLSM ADPNIRFLDKLPQQTGDRAGIKDRVYSN (SEQ ID NO:572); SIYELLENGQRAGT CVLEYATPLQTLFAMSQYSQAGFSGEDRLEQ (SEQ ID NO:573); AKLFCRTLE DILADAPESQNNCRLLIAYQEPADDSSFSLSQEVLRHLRQEEKEEVTGSLKTSAPV PSTSTMSQEPPELLISGMEKPLPLRTDFS (SEQ ID NO:574); and/or LLGLKGLA PAEISAVCEKGNFNVAHGLAWSYYIGYLRILPEL (SEQ ID NO:575).

35 Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in prostate BPH and to a lesser extent in bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, benign prostatic hypertrophy or prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male urinary system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 331 as residues: Ile-60 to Asn-69, Leu-106 to Asp-112, Glu-130 to Gly-136, Phe-160 to Glu-167, Pro-184 to Cys-190, Glu-197 to Ser-202, Arg-215 to Glu-221, Thr-237 to Pro-242.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of benign prostatic hypertrophy or prostate cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 99

This gene is expressed primarily in salivary gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders or injuries of the salivary gland. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of glandular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of disorders of, or injuries to the salivary gland or other glandular tissue.

FEATURES OF PROTEIN ENCODED BY GENE NO: 100

This gene maps to chromosome 15, accordingly, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 15. The translation product of this gene shares sequence homology with a *C.elegans* gene of unknown function. In specific embodiments, polypeptides of the invention comprise the sequence: DPRVRLNSLTCKHIFISLTQ (SEQ ID NO:583); TMKLLKLRRNIV KLSLYRHFTN (SEQ ID NO:576); TLILAVAASIVFIWTTMKFRI (SEQ ID NO:577); VTCQSDWRELWVDDAIWRLLFSMILFVI (SEQ ID NO:578); MVLWR PSANNQRFAFSPLSEEEEEDEQ (SEQ ID NO:580); KEPMLKESFEGMKMRS TKQEPNGNSKVNKAQEDDL (SEQ ID NO:584); and/or KWVEENVPSVTDVALP ALLDSDEERMITHFERSKME (SEQ ID NO:582). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in thyroid and to a lesser extent in osteoclastoma, kidney medulla, and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, thyroid dysfunction or cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 333 as residues: Lys-107 to Leu-124, Glu-150 to Thr-159, Pro-173 to Asp-179, Ser-192 to Ser-201.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of thyroid dysfunction or cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 101

This gene maps to chromosome 16, therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 16. In specific embodiments, polypeptides of the invention comprise the sequence:

IRHELTVLRDTRPACA (SEQ ID NO:585); and/or MDFXMALIYD (SEQ ID NO:586). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in kidney cortex and to a lesser extent in adult
5 brain, corpus colosum, hippocampus, and frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to
10 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
15 fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of neurological
20 disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 102

In specific embodiments, polypeptides of the invention comprise the sequence:
MQEMMRNQDRALSNLESIPGGYNA (SEQ ID NO:587); LRRMYTDIQEPMLSA
25 AQEQF GGNPF (SEQ ID NO:588); ASLVSNNTSSGEGSQPSRTENRDPLPNPWAP
QT (SEQ ID NO:589); SQSSASSGTASTVGGTTGSTASGTSGQSTTAPNLVPGV
GASMFNTPG MQSLLQQITENPQLMQNMLSAPY (SEQ ID NO:590);
MRSMMQSLSQNPDLAAQMMLNPLFAGNPQLQEQMRQQLPTFLQQ (SEQ ID
NO:591); MQNPDTLSAMSNPRAMQALLQIQGLQTLATEAPGLIPGFTPGLG
30 ALGSTGGSSGTNGSNATPSENTSPTAGT (SEQ ID NO:592); TEPGHQQFI
QQMLQALAGVNPQLQNPEVRFQQQLEQLSAMGFLNREANLQALIATGGDINAA
IERLLGSQPS (SEQ ID NO:593); RNPAMMQEMMRNQDRALSNLESIPGGY
NALRRMYTDIQEPMLSAA (SEQ ID NO:594); GNPFASLVSNNTSS (SEQ ID
NO:595); ENRDPLPNPWA (SEQ ID NO:595); GKILKDQDTLSQHGHD (SEQ ID
NO:597); GLTVHLVIKTQNRP (SEQ ID NO:598); SELQSQMQRQLLSNPEMM
35 (SEQ ID NO:599); PEISHMLNPDIMR (SEQ ID NO:600); and/or
RQLIMANPQMQLIQRNP (SEQ ID NO:601). Polynucleotides encoding these

polypeptides are also encompassed by the invention.

This gene is expressed primarily in breast.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of tumor systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of some types of breast cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 103

The translation product of this gene shares sequence homology with secreted serine proteases and lysozyme C precursor, which is thought to be important in bacteriolytic function. In specific embodiments, polypeptides of the invention comprise the sequence: NLCHVDCQDLLNP NLLAGIHCAKRIVS (SEQ ID NO:602); LDGFEGYSLSDWLCLAFVESKFN (SEQ ID NO:603); NENADGSFDYGLFQINSHYWCN (SEQ ID NO:604); and/or NLCHVDCQDLLNP NLLAGIHCAKRIVS (SEQ ID NO:605). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, infection. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 336 as residues: Ile-62 to Phe-70, Asn-78 to Asn-84.

The tissue distribution and homology to lysozyme C precursor indicates that polynucleotides and polypeptides corresponding to this gene are useful for boosting the monocyte-macrophage system and enhance the activity of immunoagents.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 104

This gene is expressed primarily in apoptotic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of some immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 105

The translation product of this gene shares sequence homology with ARI protein of Drosophila (accession: 2058299; EMBL: locus DMARIADNE, accession X98309), which is thought to be important in axonal path-finding in the central nervous system. In specific embodiments, polypeptides of the invention comprise the sequence IREVNEVIQNPAT (SEQ ID NO:606); ITRILLSHFNWDKEKLMERYF DGNLEKLFA (SEQ ID NO:607); NTRSSAQDMPCQICYLNYPNSYF (SEQ ID NO:608); TGLECGHKFCMQCWSEYLTTKIMEEGMGQTISCPAHG (SEQ ID NO:614); CDILVDDNTVMRLITDSKVKLKYQHLITNSFVECNRLWKWCPAPD CHHVVKVQYPDAKPV (SEQ ID NO:609); CDILVDDNTVMRLITDSK

VKLKYQHLITNSFVECNRLWKWCPAPDCHHVVKV (SEQ ID NO:610);
 GCNHMVCRNQNCKAFCWVCLGPWEPHGSAWYNCNRYNEDDAKAARDAQE
 RSRAALQRYL (SEQ ID NO:611); FYCNRYMNHMQSLRFEHKLYAQVKQ
 KMEEMQQHNMSWIEVQFLKKAVDVLCQCRATLMT (SEQ ID NO: 612);
 5 YVFAFYLLKNNQSIIFENNQADLENATEVLSGYLERDISQDSLQDIKQKVQDKY
 RYCESR (SEQ ID NO:613) Polynucleotides encoding these polypeptides are also
 encompassed by the invention.

This gene is expressed primarily in adult brain, and to a lesser extent in
 endometrial tumor, melanocytes, and infant brain.

10 Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, diseases or injuries involving axonal path development. Similarly,
 polypeptides and antibodies directed to these polypeptides are useful in providing
 15 immunological probes for differential identification of the tissue(s) or cell type(s). For
 a number of disorders of the above tissues or cells, particularly of the central nervous
 system, expression of this gene at significantly higher or lower levels may be routinely
 detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g.,
 serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
 20 taken from an individual having such a disorder, relative to the standard gene
 expression level, i.e., the expression level in healthy tissue or bodily fluid from an
 individual not having the disorder.

The tissue distribution and homology to ARI protein indicates that
 polynucleotides and polypeptides corresponding to this gene are useful for treatment of
 25 disease states or injuries involving axonal path development, including
 neurodegenerative diseases and nerve injury.

FEATURES OF PROTEIN ENCODED BY GENE NO: 106

The translation product of this gene shares sequence homology with cytochrome
 30 b561 [Sus scrofa] which is thought to be an integral membrane protein of
 neuroendocrine storage vesicles of neurotransmitters and peptide hormones.

This gene is expressed primarily in frontal cortex and to a lesser extent in
 rhabdomyosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as
 35 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to

these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 339 as residues: Ser-18 to Pro-24.

The tissue distribution and homology to cytochrome b561 [*Sus scrofa*] indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of neurological disorders. This gene may also be important in regulation of some types of cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 107

In specific embodiments, polypeptides of the invention comprise the sequence: MWGYLFVDAAWNFLGCLICGW (SEQ ID NO:615); MHFISSGNVSAIRSSILL RXSLSYLGNCRLRVSAIFVYFLLFLLS (SEQ ID NO:616); and/or MDQALRGSPSE GFSTDPSPPQVGRQIPSFPPWRRRLVLPKASGCFLEREWLVCVFKLRTRPGAHA HAYNSSILGGRGKGIT (SEQ ID NO:617). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in pancreas tumor and to a lesser extent in cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pancreatic tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred

epitopes include those comprising a sequence shown in SEQ ID NO: 340 as residues: Pro-22 to Phe-33.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of pancreatic tumors.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 108

This gene maps to chromosome 17 and therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 17. In specific embodiments, polypeptides of the invention comprise the sequence:

- 10 MLPALASCCHFSPPEQAARLKKLQEQEKQKVEFRKRMEKEVSDFIQDSGQIK
KKFQPMNKIERSILHDVVEVAGLTSFSFGEDDDCRYVMIFKKEFAPSDEELDSY
RRGEEWDPQKAEEKRNXXKELAQRQ (SEQ ID NO:618); EEEAAQQGPVVV
SPASDYKDKYSHLIGKGAAKDAAHMLQANKTYGCXPVANKRDTRSIEEAMNE
IRAKKRLRQSGE (SEQ ID NO:619); PPRRPAQLPLTPGAGQGAGRDKAAAIRA
15 HPGAPPLNHLLP (SEQ IDNO:620); AVPQAGGKQVFDLSPLELGYVRGMCVCV
(SEQ ID NO:621) and/or MLPALASCCHFSPPEQAARLKKLQEQEKQKVEFRK
RMEKEVSDFIQDSGQIKKKFQPMNKIERSILHDVVEVAGLTSFSFGEDDDCRYV
MIFKKEFAPSDEELDSYRRGEEWDPQKAEEKRNXXKELAQRQEEEEAAQQGPVVV
SPASDYKDKYSHLIGKGAAKDAAHMLQANKTYGCXPVANKRDTRSIEEAMNE
20 IRAKKRLRQSGE (SEQ ID NO:622). Polynucleotides encoding these polypeptides
are also encompassed by the invention. The translation product of this gene shares
sequence homology with FSA-1 which may play a role as a structural protein
component of the acrosome.

This gene is expressed primarily in fetal kidney and sperm.

25

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, male reproductive disorders, especially involving acrosomal disfunction.
Similarly, polypeptides and antibodies directed to these polypeptides are useful in
30 providing immunological probes for differential identification of the tissue(s) or cell
type(s). For a number of disorders of the above tissues or cells, particularly of the male
reproductive system, expression of this gene at significantly higher or lower levels may
be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily
fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or
35 cell sample taken from an individual having such a disorder, relative to the standard
gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an

individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 341 as residues: Glu-8 to Asn-35.

The tissue distribution and homology to FSA-1 indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of infertility due to acrosomal disfunction of sperm.

FEATURES OF PROTEIN ENCODED BY GENE NO: 109

This gene is expressed primarily in pituitary and to a lesser extent in epididymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 342 as residues: Met-1 to Trp-6.

Because the gene is found in both pituitary and epididymus, this indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of male reproductive disorders. This may involve a secreted peptide produced in the pituitary targeting the epididymus.

FEATURES OF PROTEIN ENCODED BY GENE NO: 110

In specific embodiments, polypeptides of the invention comprise the sequence: LLCPVLNSGXSWNFPHPHQPEYSFHGFHSTRLWI (SEQ ID NO:623); and/or PSTPWFLFLLGLTCPFSTSHPRWDSIPP (SEQ ID NO:624). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in resting T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, T-cell disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of certain immune disorders, especially those involving T-cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 111

This gene is expressed primarily in cerebellum and whole brain and to a lesser extent in infant brain and fetal kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 344 as residues: Asp-48 to Gly-55.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 112

The translation product of this gene shares sequence homology with yeast mitochondrial ribosomal protein homologous to ribosomal protein s15 of E.coli which

is thought to be important in the early assembly of ribosomes (See Accession No. M38016). This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in developmental tissues.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, development of cancers and tumors in addition to healing wounds. Similarly, polypeptides and antibodies directed to these polypeptides are useful in
10 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and developmental expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
15 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ribosomal protein s15 of *E. coli* indicates that polynucleotides and polypeptides corresponding to this gene are useful for
20 diseases related to the assembly of ribosomes in the mitochondria which is important in the translation of RNA into protein. Therefore, this indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of multiple tumors as well as in healing wounds which are thought to be under similar regulation as developmental tissues. Protein, as well as, antibodies directed against the
25 protein have utility as tumor markers, in addition to immunotherapy targets, for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 113

The translation product of this gene shares sequence homology with human
30 poliovirus receptor precursors which are thought to be important in viral binding and uptake. Preferred polypeptide fragments comprise the following amino acid sequence:
ELISISNVALADEGEYTCSTFTMPVRTAKSLVTVLGIPQKPIITGYKSSLREKDT
ATLNCQSSGSKPAARLTWRKGDQELHGEPTRIQEDPNGKTFTVSSSVTFQVTR
EDDGASIVCSVNHESLKGADRSTSQRIEVLYTPTAMIRPDPPHPREGQKLLHLC
35 EGRGNPVPQQYLWEKEGSVPPLKMTQESALIFPFLNKSDSGTYGCTATSNMGS
YKAYYTLNVND (SEQ ID NO:625). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. gnllPID1002627).

This gene is expressed almost exclusively in human brain tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, susceptibility to viral disease and diseases of the CNS especially cancers of that system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 346 as residues: Leu-26 to Asp-37, Lys-53 to Ser-59.

The tissue distribution and homology to poliovirus receptor precursors indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and prevention of diseases that involve the binding and uptake of virus particles for infection. It might also be helpful in genetic therapy where the goal is to insert foreign DNA into infected cells. With the help of this protein, the binding and uptake of this foreign DNA might be aided. In addition, it is expected that over expression of this gene will indicate abnormalities involving the CNS, particularly cancers of that system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 114

The translation product of this gene shares sequence homology with YO87_CAEEL hypothetical 28.5 KD protein ZK1236.7 in chromosome III of *Caenorhabditis elegans* in addition to alpha-1 collagen type III (See Accession No. gil537432). One embodiment for this gene is the polypeptide fragment(s) comprising the following amino acid sequence: VP_{ELPDRVHQLHQA}VQGCALGRPGFPGGPTH SGHHKSHPGPAGGDYNRCDRPGQVHLHNPRGTGRRGQLHPTAGPGVHRRA CPSQQLPHRLGPGVPCPSLTPVLPSWTQSWCG LPGYTSSS (SEQ ID NO:630). An additional embodiment is the polynucleotide fragment(s) encoding these polypeptide fragments

This gene is expressed primarily in brain cells and to a lesser extent in activated B and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegeneration and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 347 as residues: Glu-34 to Glu-39, Gly-51 to Ser-72, Ala-88 to Glu-93, Gln-100 to Val-105.

The tissue distribution and homology to YO87_CAEEL hypothetical 28.5 KD protein ZK1236.7 in chromosome III of *Caenorhabditis elegans* as well as to a conserved alpha-1 collagen type III protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons' Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorders. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 115

The translation product of this gene shares sequence homology with alpha 3 type IX collagen which is thought to be important in hyaline cartilage formation via its ability to uptake inorganic sulfate by cells (See Accession No. gil975657). One embodiment of this gene is the polypeptide fragment comprising the following amino acid sequence: SLRRPRSAAXQTLTTLSSVSSASSALPGSREPCDPRAPPPR SGSAASCCSCCSCPRRRAPLRSPRGSKRRIRQREVVDLYNGMCLQGPGVPG RDGSPGANGIPGTPGIPGRDGFKEGKECLRESFEESWTPNYKQCSWSSLNY GIDLGKIAECTFTKMRSNSALRVLFSGSLRLKCRNACCQRWYFTFNGAECSGP LPIEAIYLDQGSPEMNSTINIHRSSVEGLCEGIGAGLVDVAIWVGTCSDYPKG DASTGWNSVSRIIIIEELPK (SEQ ID NO:634). An additional embodiment are the

polynucleotide fragments encoding this polypeptide fragment.

This gene is expressed primarily in smooth muscle and to a lesser extent in synovial tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias i.e., spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid and autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to alpha 3 type IX collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of diseases associated with the mutation in this gene which leads to the many different types of chondrodysplasias. By the use of this product, the abnormal growth and development of bones of the limbs and spine could be routinely detected or treated in utero since the protein or muteins thereof could affect epithelial cells early in development and later the chondrocytes of the developing craniofacial structure.

FEATURES OF PROTEIN ENCODED BY GENE NO: 116

The translation product of this gene shares sequence homology with retrovirus-related reverse transcriptase which is thought to be important in viral replication. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: TKKENC RPASLMNIDTKILNKILMNQ (SEQ ID NO:640). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments (See Accession No. pirlA25313IGNHUL1).

This gene is expressed primarily in human meningima.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, retroviral diseases such as AIDS, and possibly certain cancers due to transactivation of latent cell division genes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to retrovirus-related reverse transcriptase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of diseases and maladies associated with retroviral infection since a functional reverse transcriptase (RT) or RT-like molecule is an integral component of the retroviral life cycle.

FEATURES OF PROTEIN ENCODED BY GENE NO: 117

The translation product of this gene shares sequence homology with an unknown gene from *C. elegans*, as well as weak homolog with mammalian metaxin, a gene contiguous to both thrombospondin 3 and glucocerebrosidase, is known to be required for embryonic development. Preferred polypeptide fragments comprise the following amino acid sequence: MCNLPKVVCRANAEYMSPSGKVPXXHVGNGQ VVSELGPVQFVKAKGHSLSDGLLEEVQKAEMKAYMELVNNMLLTAEYLQWC DEATVGXITHXRYGSPYPWPLXHILAYQKQWEVKRKXKAIGWGKKTLDQVLE DVDQCCQALSQRLGTQPYFFNKQPTELDALVFGHLYTILTTQLTNDELSEKVK NYSNLLAFCRRI EQHYFEDRGKGRLS (SEQ ID NO:641); MCNLPKVVCRANAE YMSPSGKVPXXHVGNGQVVSELGPVQFVK (SEQ ID NO:642). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. gil1326108).

This gene is expressed primarily in fetal tissues and to a lesser extent in hematopoietic cells and tissues, including spleen, monocytes, and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer, lymphoproliferative disorders; inflammation; chondrosarcoma, and Gaucher disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification

of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and embryonic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression in embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation or cellular division. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and survival of hematopoietic cell lineages. Thus, this gene may be useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 118

The translation product of this gene shares sequence homology with reverse transcriptase which is important in the synthesis of a cDNA chain from an RNA molecule, and is a method whereby the infecting RNA chains of retroviruses are transcribed into their DNA complements. One embodiment for this gene is the polypeptide fragment comprising the following amino acid sequence:
 MXXXNSHITIFTLNVNGLNAPNERHRLANWISQDQVCCIQETHLTGRDTHRL
 KIKGWRKIYQANGKQKK (SEQ ID NO:647). An additional embodiment is the polynucleotide fragments comprising polynucleotides encoding these polypeptide fragments (See Accession No. gil2072964).

This gene is expressed primarily in skin and to a lesser extent in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, hematopoietic disorders; inflammation; disorders of immune surveillance. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the epidermis and/or hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and

wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution and homology to reverse transcriptase indicates that polynucleotides and polypeptides corresponding to this gene are useful for cancer therapy. Expression in the skin also indicates that this gene is useful in wound healing and fibrosis. Expression by neutrophils also indicates that this gene product plays a role in inflammation and the control of immune surveillance (i.e. recognition of viral
10 pathogens). Reverse transcriptase family members are also useful in the detection and treatment of AIDS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 119

15 The translation product of this gene shares sequence homology with reverse transcriptase which is important in the synthesis of a cDNA copy of an RNA molecule, and is a method whereby a retrovirus reverse-transcribes its genome into an inheritable DNA copy.

 This gene is expressed primarily in the frontal cortex of brain.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and neurodegenerative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders
25 of the above tissues or cells, particularly of the CNS and peripheral nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,
30 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

 The tissue distribution and homology to reverse transcriptase suggest that this is useful in the treatment of cancer and AIDS. The expression in brain indicates that it plays a role in neurodegenerative disorders and in neural degeneration.

35

FEATURES OF PROTEIN ENCODED BY GENE NO: 120

One embodiment of this gene has homology to a hypothetical protein in *Schizosaccharomyces pombe* (See Accession No. 2281980). Another embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

5 IYHLHSWIFFHFKRAFCMCFITMKVIAHCSKLRKCXNAQISVFCTTLTASYPT
(SEQ ID NO:651). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments. This gene maps to chromosome 18, and therefore, may be used as a marker in linkage analysis for chromosome 18.

10 This gene is expressed primarily in adult hypothalamus and to a lesser extent in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative disorders; endocrine function; and vertigo. Similarly,
15 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, CNS and peripheral nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
20 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
25 corresponding to this gene are useful for the treatment and diagnosis of neurodegenerative disorders; diagnosis of tumors of a brain or neuronal origin; treatments involving hormonal control of the entire body and of homeostasis, behavioral disorders, such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and
30 panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo.

FEATURES OF PROTEIN ENCODED BY GENE NO: 121

35 The translation product of this gene shares sequence homology with the human IRLB protein which is thought to be important in binding to a c-myc promoter element and thus regulating its transcription (See Accession No. gil33969). This gene maps to

chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in brain and breast and to a lesser extent in a variety of hematopoietic tissues and cells.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer of the brain and breast; lymphoproliferative disorders; neurodegenerative diseases. Similarly, polypeptides and antibodies directed to these
10 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS, breast, and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
15 fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of cancer of the
20 brain, breast, and hematopoietic system. In addition, it may be useful for the treatment of neurodegenerative disorders, as well as disorders of the hematopoietic system, including defects in immune competency and inflammation. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 122

The translation product of this gene shares sequence homology with an ATP synthase, a key component of the proton channel that is thought to be important in the translocation of protons across the membrane.

30 This gene is expressed primarily in T cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, T cell lymphoma. Similarly, polypeptides and antibodies directed to these
35 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or

lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ATP synthase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of defects in proton transport, homeostasis, and metabolism, as well as the diagnosis and treatment of lymphoma. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia

FEATURES OF PROTEIN ENCODED BY GENE NO: 123

This gene maps to chromosome 15, and therefore, may be used as a marker in linkage analysis for chromosome 15.

This gene is expressed primarily in a variety of fetal tissues, including fetal liver, lung, and spleen, and to a lesser extent in a variety of blood cells, including eosinophils and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer (abnormal cell proliferation); T cell lymphomas; and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetus and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions involving cell proliferation. Expression of this gene in fetal tissues, as well as in a variety of blood cell lineages indicates that it may play a role in either cellular proliferation; apoptosis; or cell survival. Thus it may be useful in the management and

treatment of a variety of cancers and malignancies. In addition, its expression in blood cells suggest that it may play additional roles in hematopoietic disorders and conditions, and could be useful in treating diseases involving autoimmunity, immune modulation, immune surveillance, and inflammation..

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 124

This gene is expressed primarily in placenta and to a lesser extent in pineal gland and rhabdomyosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental, endocrine, and female reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the [insert system where a related disease state is likely, e.g., immune], expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 357 as residues: Leu-69 to Val-76.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders in development. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 125

This gene is expressed primarily in benign prostatic hyperplasia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of benign prostatic hyperplasia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive

system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of benign prostatic hyperplasia. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 126

This gene is expressed primarily in apoptotic T-cells and to a lesser extent in suppressor T cells and ulcerative colitis.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases involving premature apoptosis, and immunological and gastrointestinal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders involving inappropriate levels of apoptosis, especially in immune cell lineages. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases (such as AIDS), and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 127

This gene is expressed primarily in Raji cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and T cell autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 360 as residues: Asp-23 to Gly-29.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammation and T cell autoimmune disorders. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases (such as AIDS), and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 128

The translation product of this gene shares sequence homology with an *C. elegans* coding region C47D12.2 of unknown function (See Accession No. gnllPIDe348986). One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: EDDGFNRSIHEVILKNITWY SERVLTEISLGSLLILVVIRTIQYNMTRTRDKYLHTNCLAALANMSAQFRSLHQY AAQRIISLFSLLSKKHNKVLEQATQSLRGSLSNDVPLPDYAQDLNVIEEVIRMM LEIINSCLTNSLHHNPNLVALLYKRDLEQFRTHPSFQDIMQNIDLVISFFSSRLL QAGS (SEQ ID NO:657); EDDGFNRSIHEVILKNITWYSERVLTEISLGSLLILVV (SEQ ID NO:658); RTIQYNMTRTRDKYLHTNCLAALANMSAQFRSLHQYAAQ RIISLFSLLSKKHN (SEQ ID NO:659); KKHNVLEQATQSLRGSLSNDVPLPDY AQD (SEQ ID NO:661); SCLTNSLHHNPNLVYALLYKRDLEQFRTHPSFQD IMQNIDLVISFFSSRLLQAGS (SEQ ID NO:660). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments. This gene maps to

chromosome 18, and therefore, may be used as a marker in linkage analysis for chromosome 18.

This gene is expressed primarily in smooth muscle and to a lesser extent in fetal liver.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, atherosclerosis and other cardiovascular and hepatic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
10 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
15 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of circulatory system
20 disorders such as atherosclerosis, hypertension, and thrombosis. In addition, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the
25 expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 129

30 The translation product of this gene shares sequence homology with a ribosomal protein which is thought to be important in cellular metabolism, in addition to the *C.elegans* protein F40F11.1 which does not have a known function at the current time (See Accession No. gnllPIDle244552). Preferred polypeptide fragments comprise the following amino acid sequence:

35 MADIQTERAYQKQPTIFQNKKRVLGETGKEKLPRVTNKNIGLGFKDT
PRLLRGTYIDKKCPFTGNVSIRGRILSGVVTQDEDAEDHCHPPRLSALHPQVQ
PLREAPQEHVCTPVPL LQGRPDR (SEQ ID NO:662); MKMQRTTIVIRRDYLH

YIRKYNRFEKRHKNSVHLSPCFRDVQIGDIVTVGECRPLSKTVRFNVLKVTK
 AAGTKKQFQKF (SEQ ID NO:663); MADIQTERAYQKQPTIFQNKKRVLGET
 GK (SEQ ID NO:664); HCHPPRLSALHPQVQPLREAPQEHVCTPVPL LQGRPDR
 (SEQ ID NO:666); NIGLGFKDTPRLLRGTYIDKKCPFTGNVSIRGRILSGVVTQ
 5 (SEQ ID NO:669); MKMQRTTVIRRDYLHYIRKYNRFEKRHKNSVHLS (SEQ
 ID NO:667); CFRDVQIGDIVTVGECRPLSKTVRFNVLKVTKAAGTKKQFQKF
 (SEQ ID NO:668). Also preferred are polynucleotide fragments encoding these
 polypeptide fragments.

10 This gene is expressed primarily in Wilm's tumor and to a lesser extent in
 thymus and stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, diseases affecting RNA translation. Similarly, polypeptides and
 15 antibodies directed to these polypeptides are useful in providing immunological probes
 for differential identification of the tissue(s) or cell type(s). For a number of disorders
 of the above tissues or cells, particularly of the Wilm's tumors, expression of this gene
 at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
 20 fluid or spinal fluid) or another tissue or cell sample taken from an individual having
 such a disorder, relative to the standard gene expression level, i.e., the expression level
 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred
 epitopes include those comprising a sequence shown in SEQ ID NO: 362 as residues:
 Thr-11 to Asp-20.

25 The tissue distribution and homology to a ribosomal protein indicates that
 polynucleotides and polypeptides corresponding to this gene are useful for diseases
 affecting RNA translation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 130

30 The translation product of this gene shares sequence homology with a yeast
 DNA helicase which is thought to be important in global transcriptional regulation (See
 Accession No. gnlIPIDe243594). One embodiment for this gene is the polypeptide
 fragments comprising the following amino acid sequence: IFYDSDWNPTVDQQA
 MDRAHRLGQTKQVTVYRLICKGTIEERILQRAKEKSEIQRMVISG (SEQ ID
 35 NO:670); TRMIDLLEEYMVYRKHTYXRLDGSSKISERRDMVADFQNRNDI
 FVFLSTRAGGLGINLTAXDTVHF (SEQ ID NO:671); TRMIDLLEEYMVYRK
 HTYXRLDGSSKISERRDM (SEQ ID NO:674); RRDMAADFQNRNDIFVFL

STRAGGLGINLTAXDTVHF (SEQ ID NO:675) , IFYSDWNPTVDQQAMD
RAHRLGQTKQVTYRLICKG (SEQ ID NO:676); RLICKGTIEERILQRAK
EKSEIQRMVISG (SEQ ID NO:678). An additional embodiment is the polynucleotide
fragments encoding these polypeptide fragments.

5 This gene is expressed primarily in amygdala.

 Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, diseases and disorders of the brain. Similarly, polypeptides and
10 antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the central nervous system, expression of
this gene at significantly higher or lower levels may be routinely detected in certain
tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
15 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

 The tissue distribution and homology to a DNA helicase indicates that
20 polynucleotides and polypeptides corresponding to this gene are useful for diseases
affecting RNA transcription, particularly developmental disorders and healing wounds
since the later are though to approximate developmental transcriptional regulation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 131

25 This gene is expressed primarily in prostate and to a lesser extent in amygdala
and pancreatic tumors.

 Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
30 not limited to, prostate enlargement and gastrointestinal disorders, particularly of the
pancreas and gall bladder. Similarly, polypeptides and antibodies directed to these
polypeptides are useful in providing immunological probes for differential identification
of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,
particularly of the reproductive system, expression of this gene at significantly higher or
35 lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
another tissue or cell sample taken from an individual having such a disorder, relative to

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of prostate diseases, including benign prostatic hyperplasia and prostate cancer. In addition, the tissue distribution in tumors of the pancreas indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tissues where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 132

This gene is expressed primarily in adult lung and to a lesser extent in hypothalamus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pulmonary diseases and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary and respiratory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of pulmonary and respiratory disorders such as emphysema, pneumonia, and pulmonary edema and emboli. In addition, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental

disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 133

5 This gene is expressed primarily in human liver.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cirrhosis of the liver and other hepatic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of liver disorders such as cirrhosis, jaundice, and Hepatitis. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 134

 This gene is expressed primarily in fetal kidney and to a lesser extent in fetal liver and spleen.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, development and regeneration of liver and kidney and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive and excretory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 367 as residues: Pro-70 to Arg-77, Tyr-102 to Thr-107.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the kidney and liver, such as cirrhosis, kidney failure, kidney stones, and liver failure, hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells. In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 135**

This gene is expressed primarily in brain, bone marrow, and to a lesser extent in placenta, T cell, testis and neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative and immunological diseases and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 368 as residues: Met-1 to His-6.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also

play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, or sexually-linked disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 136

- 5 Translation product of this gene is homologous to the human WD repeat protein HAN11. Preferred polypeptide fragments comprise the following amino acid sequence:
- MSLHGKRKEIYKYEAPWTVYAMNWSVRPDKRFRLALGSFVEEYNNKVQLVG
LDEESSEFICRNTFDHPYPTTKLMWIPDTKGVYPDLLATSGDYLRVWRVGETET
10 RLECLLNNNKNNSDFCAPLTSFDWNEVDPYLLGTSSIDTTCTIWGLETGQVLGRV
NLVSGHVKTQLIAHDKEVYDIAFSRAGGGRDMFASVGADGSVRMFDLRHLEH
STIYEDPQHHPDLLRLCWNKQDPNYLATMAMDGMEVVILDVRVPAHLXPGTTIE
HVSMAILGPHIHPATSALQRM TTRLSSGTSSKCPEPLRTL SWPTQLXGEINNVQ
WASTQPELSPSATTAWRYSECSVGGAVPTRQGLLYFLPLPHPQS (SEQ ID
15 NO:679); MSLHGKRKEIYKYEAPWTVYAMNWSVRPDKRFRLALGSFV
EEYNNKVQLVGLDEESSEFICRNTFDHPYPTTKLMWIPDTKGVYPDLLATSGDY
LRVWRVGETETRLECLLNNNKNNSDFCAPLTSFDWNEVDPYLL (SEQ ID
NO:680); SFDWNEVDPYLLGTSSIDTTCTIWGLETGQVLGRVNLVSGHVK
TQLIAHDKEVYDIAFSRAGGGRDMFASVGADGSVRMFDLRHLEHSTIYEDPQH
20 HPLLRLCWNKQDPNYLATMAMDGMEVVILDVRVPAHLXPGTTI (SEQ ID
NO:681); VGADGSVRMFDLRHLEHSTIYEDPQHHPDLLRLCWNKQDPNYLA
TMAMDGMEVVILDVRVPAHLXPGTTIEHVSMAILGPHIHPATSALQRM TTRL
SGTSSKCPEPLRTL SWPTQLXGEINNVQWASTQPELSPSATTAWRYSECSV
GAVPTRQGLLYFLPLPHPQS (SEQ ID NO:682). Also preferred are polynucleotide
25 fragments encoding these polypeptide fragments.

This gene is expressed primarily in placenta, embryo, T cell and fetal lung and to a lesser extent in endothelial, tonsil and bone marrow.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
30 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological and developmental diseases in addition to cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the
35 immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or

cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 369 as residues: Gly-19 to Gln-28, Pro-36 to Phe-42.

5 The tissue distribution in tumors of colon, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above
10 listed tumors and tissues. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 137**

 This gene is expressed primarily in TNF and INF induced epithelial cells, T cells and kidney.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory conditions particularly inflammatory reactions in the kidney. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of renal
25 system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
30 individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 370 as residues: Thr-67 to Gly-72, Gln-132 to Ala-145, Arg-150 to Pro-157.

 The tissue distribution indicates that the protein products of this gene are useful for treating the damage caused by inflammation of the kidney.

35

FEATURES OF PROTEIN ENCODED BY GENE NO: 138

This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1 (See Accession No. D63485).

5 This gene is expressed primarily in breast cancer and colon cancer and to a lesser extent in thymus and fetal spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers, especially of the breast and colon tissues. Similarly,
10 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
15 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of colon and breast origins indicates that
20 polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 139

This gene maps to chromosome 17, and therefore, can be used as a marker for linkage analysis from chromosome 17.

This gene is expressed primarily in CD34 positive cells, and to lesser extent in activated T-cells and neutrophils.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunologically related diseases and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
35 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and hematopoietic system, expression of this gene at significantly higher or lower levels

may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in CD34, T-cell and neutrophils indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of hematopoietic disorders and immunologically related diseases, such as anemia, leukemia, inflammation, infection, allergy, immunodeficiency disorders, arthritis, asthma, immune deficiency diseases such as AIDS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 140

This gene was recently cloned by another group, who called the gene KIAA0313 gene. (See Accession No. d1021609.) Preferred polypeptide fragments comprise the amino acid sequence:

LYATATVISSPSTEXLSQDQGDRA SLDAADSGRGSWTSCSSGSHDNIQTIQ
 HQRSWETLPFGHTHFDYSGDPAGLWASSSHMDQIMFSDHSTKYNRQNSRES
 LEQAQSRASWASSTGYWGEDSEGDTGTIKRRGGKDV SIEAESSSLTSVTTEETK
 PVPMPAHIAVASSTTKGLIARKEGRYREPPPTPPGYIGIPITDFPEGHSH PARKP
 PDYNVALQR SRMVARSSDTAGPSSVQQPHGHPTSSRPVNKPQWHKXNESDPRL
 APYQSQGFSTEEDEDEQVSAV (SEQ ID NO:683); HMDQIMFSDHSTKYNRQ
 NSRESLEQAQSRASWASSTGYWGE (SEQ ID NO:684); SVTTEETKPVPMP
 AHIAVASSTTKGLIARKEGRYREPPPTPPGYIGIPITD (SEQ ID NO:685); and
 VALQR SRMVARSSDTAGPSSVQQPHGHPTSSRPVNKPQW
 HKXNESDPRLAPYQSQGF (SEQ ID NO:686). Also preferred are polynucleotide
 fragments encoding these polypeptide fragments. This gene maps to chromosome 4,
 and therefore, may be used as a marker in linkage analysis for chromosome 4 (See
 Accession No. AB002311).

This gene is expressed primarily in ovarian cancer, tumors of the Testis, brain, and colon.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, ovarian, testicle, brain and colon cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male and female reproductive systems,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of colon, ovary, testis, and brain origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 141

This gene is expressed primarily in spleen and colon cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, colon cancer and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the gastrointestinal tract and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of colon, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 142

Translation product is homologous to T cell translocation protein, a putative zinc finger factor (See Accession No. 340454), as well as to the G-protein coupled receptor TM5 consensus polypeptide (See Accession No. R50734). Preferred polypeptide

5 fragments comprise the following amino acid sequence:

CLLFVVFVSLGMRCLFWTTIVYNVLYLKHKCN TVLLCYHLCSI (SEQ ID NO:687);
ACSKLIPAFEMVMRAKDNVYHLDCFACQLCNQRXCVGDKFFLKNNXXLCQT
DYEEGLMKEGYAPXVR (SEQ ID NO:688). Also preferred are polynucleotide
fragments encoding these polypeptide fragments.

10 This gene is expressed primarily in fetal brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders including brain cancer. Similarly, polypeptides
15 and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the Central Nervous System, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
20 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
25 corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with
30 the developing embryo.

FEATURES OF PROTEIN ENCODED BY GENE NO: 143

Translation product for this gene has significant homology to the Fas ligand, which is a cysteine-rich type II transmembrane protein/tumor necrosis factor receptor
35 homolog. Mutations within this protein have been shown to result in generalized lymphoproliferative disease leading to the development of lymphadenopathy and autoimmune disease (See Medline Article No. 94185175). Preferred polypeptide

fragments comprise the following amino acid sequence:

SALSEPGAPDRRRPCPESVPRRPDDEQWPPPTALCLDVAPLPPSS (SEQ ID NO:689). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. 473565).

5 This gene is expressed primarily in osteoblasts, lung, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoblast-related, pulmonary, neurological, and immunological
10 diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
15 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 376 as residues: Trp-33 to Thr-40, Lys-
20 45 to Ile-63.

The tissue distribution in osteoblasts, lung, and brain combined with its homology to the Fas ligand indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as,
25 antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Because the Fas ligand gene is known to be expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including asthma, immune deficiency diseases such as AIDS
30 and leukemia, and various autoimmune disorders including lupus and arthritis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 144

This gene shares sequence homology with a 21.5 KD transmembrane protein in the SEC15-SAP4 intergenic region of yeast. (See Accession No. 1723971.) Preferred
35 polypeptide fragments comprise the amino acid sequence:

AHASESGERWWACCGVRFGRLRSIEAIGRSCCHDGPGLVANRGRFRFKWAIEL
SGPGGSGRGRSDRSGSGQDSLYPVGYLDKQVPDTSVQETDRILVEKRCWDIAL

GPLKQIPMNLFI MYMAGNTISIFPTMMVCMMAWRPIQALMAISATFKMLESSSQ
 KFLQGLVYLIGNLMGLALAVYKCQSMGLLP₅THASDWLAFIEPPERMEFSGG
 GLLL (SEQ ID NO:691); PVGYLDKQVPDTSVQETDRILVEKRCWDIALGPLKQ
 IPMNLFI (SEQ ID NO:693); and ATFKMLESSSQKFLQGLVYLIGNLMGLALAV
 YKCQSMGLLP₁₀THASD (SEQ ID NO:692). Also preferred are polynucleotide
 fragments encoding these polypeptide fragments.

This gene is expressed primarily in osteoclastoma, hemangiopericytoma, liver,
 lung.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, osteoclastoma, hemangiopericytoma, liver and lung tumors. Similarly,
 polypeptides and antibodies directed to these polypeptides are useful in providing
 immunological probes for differential identification of the above tissue(s) or cell
 type(s). For a number of disorders of the above tissues or cells, particularly of the lung
 and liver systems, expression of this gene at significantly higher or lower levels may be
 routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily
 fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or
 cell sample taken from an individual having such a disorder, relative to the standard
 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
 individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
 corresponding to this gene are useful for diagnosing osteoclastoma,
 hemangiopericytoma, liver and lung tumors.

FEATURES OF PROTEIN ENCODED BY GENE NO: 145

Translation product of this gene shares homology with the glucagon-69 gene
 which may indicate this gene plays a role in regulating metabolism. (See Accession No.
 A60318) One embodiment for this gene is the polypeptide fragments comprising the
 following amino acid sequence:
 PTTKLDIMEKKKKHIQIRFPSFYHKLVD₃₀SGRMRSKRETRREDS₃₅DTKHNL (SEQ ID
 NO:694). An additional embodiment is the polynucleotide fragments encoding these
 polypeptide fragments.

This gene is expressed primarily in brain, kidney, colon, and testis.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, brain, kidney, colon, and testicular cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, neurological, circulatory, and gastrointestinal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of brain, kidney, colon, and testis origins, indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 146

The translation product of this gene shares sequence homology with goliath protein which is thought to be important in the regulation of gene expression during development. Protein may serve as a transcription factor. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

TEHIIAVMITELRGKDILSYLEKNISVQMTIavgTRMPPKNFSRGS�VFVSISFIV
LMISSAWLIFYFIQKIRYTNARDRNQRRLGDAAKKAISKLTTRTVKKGDKETD
PDFDHCAVCIESYKQNDVVRILPCKHVfHKSCVDPWLSEHCTCPMCKLNILKA
LGIV (SEQ ID NO:695); TEHIIAVMITELRGKDILSYLEKNISVQMTIavgTRMP
PKNFSRGS�VFVSISFIVLM IISSAWLIFYF (SEQ ID NO:697); SISFIVLMISSA
WLIFYFIQKIRYTNARDRNQRRLGDAAKKAISKLTTRTVKKGDKE (SEQ ID
NO:698); VKKGDKETDPDFDHCAVCIESYKQNDVVRILPCKHVfHKSCVDP

WLSEHCTCPMCKLNILKALGIV (SEQ ID NO:699). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments (See Accession No. 157535). Moreover, another embodiment is the polynucleotide fragments encoding these polypeptide fragments:

- 5 MTHPGTEHIIA VMITELRGKDILSYLEKNISVQMTIA VGTRMPPKNFSRGS
LVFVSISFIVLMISSAWLIFYFIQKIRYTNARDRNQRR LGDAAKKAISKLTTRTV
KKGDKETDPDFDHCAVCIESYKQNDVVRILPCKHVFHKSCVDPWLSEHCTCP
MCKLNILKALGIVPNLPCTDNVAFDMERLRTQAVNRRSALGDLAGDNSLGL
PLRTSGISPLPDGELTPRTGEINIA VTKEWFIIASFGLLSALTLCYMIIRATASLN
10 ANEVEWF (SEQ ID NO:696); MTHPGTEHIIA VMITELRGKDILSYLEKNISVQ
MTIAVGTRMPPKNFSRGS LVFVSISFIVLMISSAWLIFYFIQKIRYTNARDRNQRR
LGDAAKKAISKLTTRT (SEQ ID NO:700); AAKKAISKLTTRTVKKGDK
ETDPDFDHCAVCIESYKQNDVVRILPCKHVFHKSCVDPWLSEHCTCPMCKLNIL
KALGIVPNLPC (SEQ ID NO:701); TQAVNRRSALGDLAGDNSLGLPLRTSGI
15 SPLPDGELTPRTGEINIA VTKEWFIIASFGLLSALTLCYMIIRATASLNANEVEW
F (SEQ ID NO:702); PLHGVADHLGCDPQTRFFVPPNIKQWIALLRGNCTF
KEKISRAAFHNAVAVVIYNNKSKEEPVTMTHPGTEHIIA VMITELRGKDILSYLE
KNISVQMTIA VGTRMPPKNFSRGS LVFVSISFIVLMISSAWLIFYFIQKIRYTN
ARDRNQRR LGDAAKKAISKLTTRTVKKGDKETDPDFDHCAVCIESYKQNDVVR
20 LPCKHVFHKSCVDPWLSEHCTCPMCKLNILKALGIVPNLPCTDNVAFDMERLT
RTQAVNRRSALGDLAGDNSLGLPLRTSGISPLPDGELTPRTGEINIA VTKEW
FIASFGLLSALTLCYMIIRATASLNANEVEWF (SEQ ID NO:703); and
HGVADHLGCDPQTRFFVPPNIKQWIALLRGNCTFKEKISRAAFHNAVAVVIY
NNKSKEE (SEQ ID NO:704). An additional embodiment is the polynucleotide
25 fragments encoding these polypeptide fragments. When tested against Jurkat cell lines,
supernatants removed from cells containing this gene activated the GAS pathway.
Thus, it is likely that this gene activates immune cells through the JAKS/STAT signal
transduction pathway.

- 30 This gene is expressed primarily in macrophage, breast, kidney and to a lesser
extent in synovium, hypothalamus and rhabdomyosarcoma.

- Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, schizophrenia and cancer. Similarly, polypeptides and antibodies directed
35 to these polypeptides are useful in providing immunological probes for differential
identification of the tissue(s) or cell type(s). For a number of disorders of the above
tissues or cells, particularly of the immune and neural system, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to zinc finger protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of schizophrenia, kidney disease and other cancers. The tissue distribution in macrophage, breast, and kidney origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of tumors within these tissues, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 147

The translation product of this gene shares sequence homology with HNP36 protein, an equilibrative nucleoside transporter, which is thought to be important in gene transcription as well as serving as an important component of the nucleoside transport apparatus (See Accession No. 1845345). One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

MSGQGLAGFFASVAMICAIASGSELSESAFGYFITACAVIILTICYLGLPRLEFYR
 YYQQLKLEGPGEQETKLDLISKGEEPRAGKEESGVSVSNSQPTNESHSHKAILK
 NISVLAFSVCFITTTIGMFPAVTVEVKSSIAGSSTWERYFIPVSCFLT FNIFDWLG
 RSLTAVFMWPGKDSRWLPSWXLARLVFVPLLLLCNIKPRRYLTVVFEHDAWFI
 FFMAAFASNGYLASLCMCFGPKKVKPAEAETAEPSPSSCVVWWHWGLFS
 PSCSGQLCDKGWTEGLPASLPVCLLPLPSARGDPEWGGFFF (SEQ ID
 NO:705); MSGQGLAGFFASVAMICAIASGSELSESAFGYFITACAVIILTIC
 YLGLPRLEFYRYYQQLKLE GPGEQETKLDLISKGEEPRAGKEESGVSVSNSQ
 PTNESHSHI (SEQ ID NO:706); SGVSVSNSQPTNESHSHKAILKNISVLAFSVCFI
 FTTTIGMFPAVTVEVKSSIAGSSTWERYFIPVSCFLT FNIFDWLGRS (SEQ ID
 NO:707), TIGMFPAVTVEVKSSIAGSSTWERYFIPVSCFLT FNIFDWLGRSLTAVF
 MWPGKDSRWLPSWXLARLVFVPLLLLCNIK PRRYLTVVFEHDA (SEQ ID
 NO:708); FGPKKVKPAEAETAEPSPSSCVVWWHWGLFSPSCSGQLCDK

GWTEGLPASLPVCLLPSPARGDPEWSSGGFFF (SEQ ID NO:709). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in eosinophils and aortic endothelium and to a lesser extent in umbilical vein endothelial cell and thymus.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential
10 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having
15 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to HNP36 protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of blood neoplasias and other hematopoietic disease.

20 **FEATURES OF PROTEIN ENCODED BY GENE NO: 148**

This gene is expressed primarily in breast cancer cell lines, thymus stromal cells, and ovary.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, endocrine and female reproductive system diseases including breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell
30 type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard
35 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of endocrine disorders. In addition, the tissue distribution in tumors of thymus, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 149

Translation product of this gene has homology to pmt1 and pmt 2, two conserved schizosaccharomyces pombe genes. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:
 15 DDDGFEIVPIEDPAKHRILDPEGLALGAVIASSKKAKRDLIDNSFNRYTFNEDEG
 ELPEWVFVQEEKQHRIRQLPVGKKEVEHYRKRWREINARPIXXXXXXXXXXXXX
 XXXXXXLEQTRKKAEAVVNTVDIXRTRES (SEQ ID NO:710);
 DDDGFEIVPIEDPAKHRILDPEGLALGAVIASSKKAKRDLIDNSFNRYTF (SEQ
 ID NO:711); KRWREINARPIXXXXXXXXXXXXXXXXXXXXXLEQTRKKAE
 AVVNTVDIXRTRES (SEQ ID NO:712). An additional embodiment is the
 20 polynucleotide fragments encoding these polypeptide fragments (See Accession No.
 e1216734).

This gene is expressed primarily in retina and ovary and to a lesser extent in breast cancer cell, epididymus and osteosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as
 25 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, neuronal growth disorders, cancer and reproductive system disorders.
 Similarly, polypeptides and antibodies directed to these polypeptides are useful in
 providing immunological probes for differential identification of the tissue(s) or cell
 30 type(s). For a number of disorders of the above tissues or cells, particularly of the
 neural and reproductive system, expression of this gene at significantly higher or lower
 levels may be routinely detected in certain tissues (e.g., cancerous and wounded
 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
 another tissue or cell sample taken from an individual having such a disorder, relative to
 35 the standard gene expression level, i.e., the expression level in healthy tissue or bodily
 fluid from an individual not having the disorder. Preferred epitopes include those
 comprising a sequence shown in SEQ ID NO: 382 as residues: Met-1 to Gly-7.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis or treatment of reproductive system disease and cancers.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 150

One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

MIKDKGRARTALTSSQPAHLCPENPLLHLKAAVKEKKRNKKKKTIGSPKRIQS
 PLNNKLLNSPAKTLPGACGSPQKLIDGFLKHEGPPAEKPLEELSASTSGVPGLS
 10 SLQSDPAGCVRPPAPNLAGAVEFNDVKTLREWITTISDPMEEDILQVVKYCTD
 LIEEKDLEKLDLVIKYMKRLMQQSVEVWNMAFDNFILDNVQVVLQQTYGSTLK
 VT (SEQ ID NO:713); MIKDKGRARTALTSSQPAHLCPENPLLHLKAAVKE
 KKRNNKKKKTIGSPKRIQ (SEQ ID NO:714); KRIQSPLNNKLLNSPAKT
 LPGACGSPQKLIDGFLKHEGPPAEKPLEELSASTSGVPGLSSLQSDPAGCVRPP
 15 APNLAGAVEFNDVKTLREWITTISDPM (SEQ ID NO:715);
 TISDPMEEDILQVVKYCTDLIEEKDLEKLDLVIKYMKRLMQQSVE
 SVWNMAFDNFILDNVQVVLQQTYGSTLKVT (SEQ ID NO:716). An additional
 embodiment is the polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in 12 week embryo and to a lesser extent in
 20 hemangiopericytoma and frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, growth disorders and hemangiopericytoma. Similarly, polypeptides and
 25 antibodies directed to these polypeptides are useful in providing immunological probes
 for differential identification of the tissue(s) or cell type(s). For a number of disorders
 of the above tissues or cells, particularly of the circular and neural system, expression
 of this gene at significantly higher or lower levels may be routinely detected in certain
 tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
 30 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
 individual having such a disorder, relative to the standard gene expression level, i.e.,
 the expression level in healthy tissue or bodily fluid from an individual not having the
 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
 NO: 383 as residues: Leu-4 to Lys-11.

35 The tissue distribution indicates that polynucleotides and polypeptides
 corresponding to this gene are useful for the treatment of growth disorders,
 hemangiopericytoma and other soft tissue tumors.

FEATURES OF PROTEIN ENCODED BY GENE NO: 151

The translation product of this gene has been found to have homology to a human DNA mismatch repair protein PMS3. Preferred polypeptide fragments comprise the following amino acid sequence: FCHDCKFPEASPAMNCEP (SEQ ID NO:717). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. R95250).

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, lymphoma, immunodeficiency diseases, and cancers resulting from genetic instability. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 384 as residues: Met-1 to Lys-6.

The tissue distribution in neutrophils and the sequence homology indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of Hodgkin's lymphoma, since the elevated expression and secretion by the tumor mass may be indicative of tumors of this type. Additionally the gene product may be used as a target in the immunotherapy of the cancer. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. Furthermore, its homology to a known DNA repair protein would suggest gene may be useful in establishing cancer predisposition and prevention in gene therapy applications.

FEATURES OF PROTEIN ENCODED BY GENE NO: 152

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, infectious diseases and lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of inflammation and infectious diseases.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 153**

One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

MASSVPAGGHTRAGGIFLIGKLDLEASLFKSFQWLPFVLRKKC
 NFFCWDSSAHSLSLPLHPLSASCSAPACHASDTHLLYPSTRALCPSIFAWLVAPHS
 20 VFRTNAPGPTSSQSSPVFPVFPVSFIMALIVCXLVCC (SEQ ID NO:720);
 MASSVPAGGHTRAGGIFLIGKLDLEASLFKSFQWLPFVLRKKCNFFCWDSSAH
 SLPLHPLSASCSAPACHA (SEQ ID NO:721);FAWL VAPHSVFRTNAPGPTPS
 SQSSPVFPVFPVSFIMALIVCXLVCC (SEQ ID NO:722). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments.

25 This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and infectious disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred

epitopes include those comprising a sequence shown in SEQ ID NO: 386 as residues: Ser-11 to Pro-17.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of infectious diseases and inflammation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 154

This gene is expressed in multiple tissues including ovary, uterus, adipose tissue, brain, and the liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, uterine, ovarian, brain, and liver cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnostic or therapeutic uses in the treatment of the female reproductive system, obesity, and liver disorders, particularly cancer in the above tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 155

This gene maps to chromosome 3, and therefore, may be used as a marker in linkage analysis for chromosome 3 (See Accession No. D87452).

This gene is expressed in multiple tissues including brain, aortic endothelial cells, smooth muscle, pituitary, testis, melanocytes, spleen, neutrophils, and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological disorders including immunodeficiencies, cancers of the brain and the female reproductive system, as well as cardiovascular disorders, such as

atherosclerosis and stroke. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution suggest that polynucleotides and polypeptides corresponding to this gene are useful in treatment/detection of disorders in the nervous system, including schizophrenia, neurodegeneration, neoplasia, brain cancer as well as cardiovascular and female reproductive disorders including cancer within the above tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 156

The translation product of this gene shares sequence homology with the human gene encoding cytochrome b561 (See Accession No. P10897). Cytochrome b561 is a transmembrane electron transport protein that is specific to a subset of secretory vesicles containing catecholamines and amidated peptides. This protein is thought to supply reducing equivalents to the intravesicular enzymes dopamine-beta-hydroxylase and alpha-peptide amidase. Preferred polypeptides of the invention comprise the amino acid sequence:

MAMEGYWRFLALLGSALLVGFLSVIFALVWVLHYREGLGWDGSALEFNWHP
VLMVTGFVFIQGIATVYRLPWTWKCSKLLMKSIHAGLNAVAAILAISVVAVFE
NHNVNNIANMYSLSHWVGLIAVICYLLQLLSGFSVFLLPWAPLSLRAFLMPIHV
YSGIVIFGTVIATALMGLTEKLIFSLRDPAYSTFPPEGVFVNTLGLLLILVFGALIF
WIVTRPQWKRPKEPNSTILHPNGGTEQGARGSM PAYSGNNMDKSDSEL
NSEVAARKRNALALDEAGQRSTM (SEQ ID NO:724); as well as antigenic fragments

of at least 20 amino acids of this gene and/or biologically active fragments. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in anergic T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune system and metabolism related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological

probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product or RNA of this gene is useful for treatment or diagnosis of immune system and metabolic diseases or conditions including Tay-Sachs disease, phenylketonuria, galactosemia, various porphyrias, and Hurler's syndrome.

FEATURES OF PROTEIN ENCODED BY GENE NO: 157

The translation product of this gene shares sequence homology with collagen which is important in mammalian development. This gene also shows sequence homology with bcl-2. (See Accession No. P80988.) Preferred polypeptide fragments comprise the amino acid sequence: PGRAGPSPGLSLQLPAEPGHPAGNLAPLTSRPQPLCRIPAVPG (SEQ ID NO:725). Also preferred are polynucleotide sequences encoding this polypeptide fragment.

This gene is expressed primarily in HL-60 tissue culture cells and to a lesser extent in liver, breast, and uterus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological diseases, hereditary disorders involving the MHC class of immune molecules, as well as developmental disorders and reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and reproductive system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those

comprising a sequence shown in SEQ ID NO: 390 as residues: Ser-39 to Gly-46, Leu-49 to Ala-62.

- 5 The tissue distribution and homology to collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hereditary MHC disorders and particularly autoimmune disorders including rheumatoid arthritis, lupus, scleroderma, and dermatomyositis, as well as many reproductive disorders, including cancer of the uterus, and breast tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 158

- 10 This gene is expressed primarily in the amygdala region of the brain.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, a variety of brain disorders, particularly those effecting mood and personality. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and/or diagnosis of a variety of brain disorders, particularly bipolar disorder, unipolar depression, and dementia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 159

- 30 This gene is expressed in a variety of tissues and cell types including brain, smooth muscle, kidney, salivary gland and T-cells.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers of a variety of organs including brain, smooth muscle, kidney, salivary gland and T-cells and cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the central nervous, urinary, salivary, digestive, and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
5 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain, smooth muscle, and T-cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of
10 various neurological, and cardiovascular disorders, but not limited to cancer within the above tissues. Additionally the gene product may be used as a target in the immunotherapy of the cancer. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma,
15 immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 160

The translation product of this gene shares sequence homology with collagen which is thought to be important in cellular interactions, extracellular matrix formation,
20 and has been found to be an identifying determinant in autoimmune disorders. Moreover, this gene shows sequence homology with the yeast protein, Sls1p, an endoplasmic reticulum component, involved in the protein translocation process in Yeast *Yarrowia lipolytica*. (See Accession No. 1052828; see also J. Biol. Chem. 271, 11668-11675 (1996).) With mouse, this same region shows sequence homology with
25 the heavy chain of kinesin. (See Accession No. 2062607.) Recently, suppression of the heavy chain of kinesin was shown to inhibit insulin secretion from primary cultures of mouse beta-cells. (See Endocrinology 138 (5), 1979-1987 (1997).) Moreover, kinesin was found associated with drug resistance and cell immortalization. (See 468355.) Thus, it is likely that this gene also act as a genetic suppressor elements.

30 This gene is expressed primarily in the greater omentum and to a lesser extent in a variety of organs and cell types including gall bladder, stromal bone marrow cells, lymph node, liver, testes, pituitary, and thymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the endocrine, gastrointestinal, and immunological systems, including autoimmune disorders and cancers in a variety of organs and cell types.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and gastrointestinal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 393 as residues: Asn-27 to Leu-47, Gln-81 to Lys-88, Asp-93 to Lys-102, Asn-107 to Leu-116, Met-129 to Glu-141, Glu-150 to Asp-157, Lys-176 to Glu-185, Glu-333 to Tyr-349, Cys-393 to Leu-403, Gln-423 to Gly-429.

The tissue distribution in within various endocrine and immunological tissues combined with the sequence homology to a conserved collagen motif indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis of various autoimmune disorders including, but not limited to, rheumatoid arthritis, lupus erythematosus, scleroderma, dermatomyositis. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 161

This gene has homology to the tissue inhibitor of metalloproteinase 2. Such inhibitors are vital to proper regulation of metalloproteins such as collagenases (See Accession No. P16368). In addition, this gene maps to chromosome 17, and therefore, may be used as a marker in linkage analysis for chromosome 17 (See Accession No. P16368).

This gene is expressed primarily in several types of cancer including osteoclastoma, chondrosarcoma, and rhabdomyosarcoma and to a lesser extent in several non-malignant tissues including synovium, amygdala, testes, placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, various types of cancer, particularly cancers of bone and cartilage, as well as various autoimmune disorders. Similarly, polypeptides and antibodies directed

to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the musculoskeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in various cancers and the sequence homology to a collagenase inhibitor indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection of various autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 162

This gene is homologous to the mitochondrial ATP6 gene and therefore is likely a homolog of this gene family (See Accession No. X76197).

This gene is expressed primarily in brain tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, a variety of brain disorders, including Down's syndrome, depression, Schizophrenia, and epilepsy. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain tissue indicates this gene is useful for diagnosis of various neurological disorders including, but not limited to, brain cancer.

Additionally the gene product may be used as a target in the immunotherapy of cancer in the brain as well as for the diagnosis of metabolic disorders such as obesity Tay-Sachs disease, phenylketonuria and Hurler's Syndrome.

FEATURES OF PROTEIN ENCODED BY GENE NO: 163

This gene is expressed primarily in placenta, neutrophils, and microvascular endothelial cells and to a lesser extent in multiple tissues including brain, prostate, spleen, thymus, and bone.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neutropenia and other diseases of the immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in placenta indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis various female reproductive disorders. Additionally the gene product may be used as a target in the immunotherapy of various cancers. Because the gene is expressed in some cells of lymphoid and endocrine origin, the natural gene product may be involved in immune functions and metabolism regulation, respectively. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 164

This gene is expressed primarily in neutrophils, monocytes, bone marrow, and fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune system disorders including, but not limited to, autoimmune disorders such as lupus, and immunodeficiency disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having
 5 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in various immune system tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis of various immunological disorders such as Hodgkin's lymphoma, arthritis,
 10 asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 165

The translation product of this gene shares sequence homology with dystrophin which is thought to be defective in both Duchene and Becker Muscular Dystrophy.

15 Preferred polypeptide fragments comprise the following amino acid sequence:
 MKLLGECSSSIDSVKRLEHKLKEEEESLPGFVNLHSTETQTAGVIDRWELLQAQ
 ALSKELRMKQNLQKWQQFNSDLNSIWA WLGDTEEELEQLQRLELSTDIQTIELQ
 IKKLKELQKAVDHRKAIILSINLCSPEFTQADSKESRDLQDRLXQMNGRWDRV
 CSLLEEWRGLLQDALMQCQGFHEMSHGLLLMLENIDRRKNEIVPIDSNLDAEIL
 20 QDHHKQLMQIKHELLESQLRVASLQDMSCQLLVNAEGTDCLEAKEKVHVIGNR
 LKLLLKEVSRHIKELEKLLDVSSSQDLSSWSSADELDTSGSVSPXSGRSTPNR
 QKTPRGKCSLSQPGPSVSSPHSRSTKGGSDSSLSEXPGRSGRGFLFRVLRAA
 LPLQLLLLLLIGLACL VPMSEEDYSCALSNNFARSFHPMLRYTNGPPPL (SEQ ID
 NO:726); MKLLGECSSSIDSVKRLEHKLKEEEESLPGFVNLHSTETQTAGVIDR
 25 WELLQAQALSKELRMKQNLQKWQQFNSDLNSIWA WLGDTEEELEQLQRLELS
 TDIQTIELQIK (SEQ ID NO:727); KLKELQKAVDHRKAIILSINLCSPEFTQADSK
 ESRDLQDRLXQMNGRWDRVCSLLEEWRGLLQDALMQCQGFHEMSHGLLLML
 ENIDRRKNEIVPIDSNLDAEILQDHHKQLMQIKHELLESQLRVASLQDMSCQL
 (SEQ ID NO:728); QDMSCQLLVNAEGTDCLEAKEKVHVIGNRLKLLLKEVS
 30 RHIKELEKLLDVSSSQDLSSWSSADELDTSGSVSPXSGRSTPNRQKTPRGKCS
 LSQPGPSVSSPHS (SEQ ID NO:729); DSSLSEXPGRSGRGFLFRVLRAAL
 PLQLLLLLLIGLACL VPMSEEDYSCALSNNFARSFHPMLRYTNGPPPL (SEQ ID
 NO:730). Also preferred are polynucleotide fragments encoding these polypeptide
 fragments. Furthermore, this gene maps to chromosome 6, and therefore, may be used
 35 as a marker in linkage analysis for chromosome 6 (See Accession No. N62896).

This gene is expressed in numerous tissues including the heart, kidney, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, musculoskeletal disorders including Muscular Dystrophy and cardiovascular diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscle tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to dystrophin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of Muscular Dystrophy and other muscle disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 166

This gene is expressed primarily in human cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the central nervous system, including Alzheimer's Disease, Parkinson's Disease, ALS, and mental illnesses. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 399 as residues: Pro-20 to Gly-26, Leu-37 to Pro-42, His-57 to Gly-63.

The tissue distribution indicates that the protein products of this gene are useful for treatment/diagnosis of diseases of the central nervous system and may protect or

enhance survival of neuronal cells by slowing progression of neurodegenerative diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 167

- 5 Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MKLLICGNYLAPSHSESSRRCCLLCFYPLCLEINFGMKVFLSMPFLVLFQ

- SLIQED (SEQ ID NO:731). Polynucleotides encoding such polypeptides are also provided. This gene is believed to reside on chromosome 15. Therefore polynucleotides
10 derived from this gene are useful in linkage analysis as chromosome 15 markers.

This gene is expressed primarily in human testes tumor and to a lesser extent in normal human testes.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
15 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the testes, particularly cancer, and other reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of
20 the male reproductive tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily
25 fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for treatment/diagnosis of testicular diseases including cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 168

- 30 This gene is expressed primarily in fetal liver.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, conditions affecting hematopoietic development and metabolic diseases.
35 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the

hepatic system, and fetal hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 401 as residues: His-7 to Trp-17, Leu-19 to Lys-27, Pro-33 to Gly-44, Lys-68 to Gly-74, Lys-85 to Cys-95.

The tissue distribution indicates that the protein products of this gene are useful for treatment/diagnosis of diseases of the developing liver and hematopoietic system, and act as a growth differentiation factor for hematopoietic stem cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 169

The polypeptide encoded by this gene is believed to be a membrane bound receptor. The extracellular domain of which is expected to consist of the following amino acid sequence:

RILLVKYSANEENKYDYLPTTVNVCSELVKLVFCVLVSFCVIKKDHQSRNLKY
ASWKEFSDFMKWSIPAFLYFLDNLIVFYVLSYLPAMAVIFS NFSIITTALLFRIV
LKXRLNWIQWASLLTLFLSIVALTAGTKTLQHNLAGRGFHHD AFFSPSNSCLL
FRNECPRKDNCTAKEWTFPEAKWNTTARVFSHIRLGMGHVLIIVQCFISSMANI
YNEKILKEGNQLTEXIFIQNSKLYFFGILFNGLTLGLQRSNRDQIKNCGFFYGH
S (SEQ ID NO:732). Thus, preferred polypeptides encoded by this gene comprise the extracellular domain as shown above. It will be recognized, however, that deletions of either end of the extracellular domain up to the first cysteine from the N-terminus and the first cysteine of the C-terminus, is expected to retain the biological functions of the full-length extracellular domain because the cysteines are thought to be responsible for providing secondary structure to the molecule. Thus, deletions of one or more amino acids from either end (or both ends) of the extracellular domain are contemplated. Of course, further deletions including the cysteines are also contemplated as useful as such polypeptides is expected to have immunological properties such as the ability to evoke and immune response. Polynucleotides encoding all of the foregoing polypeptides are provided.

This gene is expressed primarily in human osteoclastoma and to a lesser extent in hippocampus and chondrosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, cancers, particularly those of the bone and connective tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 402 as residues: Met-1 to Cys-6, Ala-41 to Tyr-49, Lys-76 to Lys-84.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis of cancers of the bone and connective tissues, and may act as growth factors for cells involved in bone or connective tissue growth.

FEATURES OF PROTEIN ENCODED BY GENE NO: 170

Preferred polypeptides encoded by this gene comprising the following amino acid sequence:

NSVPNLQTLAVLTEAIGPEPAIPRXPREPPVATSTPATPSAGPQPLPTGTV

LVPGGPAPPCLGEAWALLPPCRPSLTSCFWSPRPSPWKETGV (SEQ ID NO:733). Polynucleotides encoding such polypeptides are also provided herein.

This gene is expressed primarily in hematopoietic progenitor cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the blood including cancer and autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the blood/circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 403 as residues: Gln-4 to His-10, Pro-25 to His-32.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis of diseases involving growth differentiation of hematopoietic cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 171

5 Preferred polypeptides encoded by this gene comprise the following amino acid sequences: ALQLAFYPDAVEEWLEENVHPSLQRLQXLLQDLSEVSAPP (SEQ ID NO:734); and/or CHPPALAGTLLRTPEGRAHARGLLLEAGGA (SEQ ID NO:735). Polynucleotides encoding such polypeptides are also provided. The protein product of this gene shares sequence homology with metallothionines. Thus, polypeptide encoded
10 by this gene are expected to have metallothionine activity, such activities are known in the art and described elsewhere herein.

This gene is expressed primarily in kidney cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
15 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the kidney including cancer and renal dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the renal system,
20 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the
25 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 404 as residues: Ser-47 to Gln-52.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases of the kidney including kidney failure.

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 172

This gene is expressed primarily in 12 week old early stage human.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for

differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing embryo, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 405 as residues: Gln-31 to Thr-43, Gly-51 to Ser-58, Pro-65 to Pro-72.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of developmental problems with fetal tissue. The gene may be involved in vital organ development in the early stage, especially hematopoiesis, cardiovascular system, and neural development.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 173**

The translation product of this gene shares sequence homology with TGN38, an integral membrane protein previously shown to be predominantly localized to the trans-Golgi network (TGN) of cells.

This gene is expressed primarily in developing embryo and to a lesser extent in cancer tissues including lymphoma, endometrial, prostate and colon.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing fetus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 406 as residues: His-65 to Ser-72, Pro-82 to Gly-91, Pro-98 to Glu-118, Ser-126 to Gly-166, Pro-180 to Asp-188, Tyr-209 to Lys-214, Gln-220 to Leu-228.

The tissue distribution and homology to an integral membrane protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for

diagnosis of cancers and developmental abnormalities where aberrant expression relates to an abnormality.

FEATURES OF PROTEIN ENCODED BY GENE NO: 174

5 The translation product of this gene shares sequence homology with a dnaJ heat shock protein from *E. coli* which is allelic to sec63, a gene that affects transit of nascent secretory proteins across the endoplasmic reticulum in yeast.

 This gene is expressed primarily in Hodgkin's lymphoma and to a lesser extent in testes.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer. Similarly, polypeptides and antibodies directed to these
15 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
20 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 407 as residues: Thr-13 to Trp-21, Arg-74 to Asp-81.

 The tissue distribution and homology to dnaJ indicates that polynucleotides and
25 polypeptides corresponding to this gene are useful as a diagnostic for cancer including Hodgkin's lymphoma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 175

30 This gene is expressed primarily in endothelial cells and to a lesser extent in bone marrow stromal cells.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases involving angiogenic abnormalities including diabetic
35 retinopathy, macular degeneration, and other diseases including arteriosclerosis and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for treating diseases where an increase or decrease in angiogenesis is indicated and as a factor in the wound healing process.

FEATURES OF PROTEIN ENCODED BY GENE NO: 176

The translation product of this gene shares sequence homology with MAT8 (mouse) which is thought to be important in regulating chloride conductance in cells (particularly in the breast) by modulating the response mediated by cAMP and protein kinase C to extracellular signals.

This gene is expressed primarily in amniotic cells and hematopoietic cells including macrophages, Neutrophils, T cells, TNF induced aortic endothelium and to a lesser extent in testes, TNF induced epithelial cells, and smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory responses mediated by T cells, macrophages, and/or neutrophils particularly those involving TNF, and also cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 409 as residues: Thr-19 to Ala-33, Leu-54 to Asp-82, Pro-89 to Ala-97, Pro-100 to Lys-125, Ser-127 to Phe-135, Gly-164 to Leu-169, Cys-173 to Arg-178.

The tissue distribution and homology to mat-8 indicates that polynucleotides and polypeptides corresponding to this gene are useful for modifying inflammatory

responses to cytokines such as TNF and thus modifying the duration and/or severity of inflammation. Polynucleotides and polypeptides derived from this gene are thought to be useful in the diagnosis and treatment of cancer.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 177

This gene is expressed primarily in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, vascular restenosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases associated with vascular response to injury such as vascular restenosis following angioplasty..

FEATURES OF PROTEIN ENCODED BY GENE NO: 178

One embodiment of the claimed invention comprises:

MRPDWKAGAGPGGPPQKPAPSSQRKPPARPSAAAAIAVAAAEERRLRQRN
RLRLEEDKPAVERCLEELVFGDVENEDALLRRLRGPRVQEHEDESGDSEVENEAKGNFPPQKKPVWVDEEDEDEEMVDMMNRRFRKDMMKNAESKLSKDNLKKRLKEEFQHAMGGVPAWAETTKRKTSSDDESEDEDDLLQRTGNFISTSTSLPRGILKMKNQCQHANARPTVARISICAVPSRCTDCDGCWD (SEQ ID NO:737); or
CLEELVFGDVENEDALLRRLRGPRVQEHEDESGDSEVENEAKGNFPPQKKPVWVDEEDEDEEMVDMMNRRFRKDMMKNAESKLSKDNLKKRLKEEFQHAMGGVPAWAETTKRKTSSDDESEDEDDLLQRTGNFISTSTSLPRGILKMKNQCQHANARPTVARISICAVPSRCTDCDGC (SEQ ID NO: 738). LKEKIVRSFEVSPDGSFLLINGIAGYLHLLAMKTKELIGSMKINGRVAASTFSSDSKKVYASSGDGEVYVWDVNSRKCLNRFVDEGSLYGLSIATSRNGQYVACGSNCGVVNIYNQDSCLQETNPKPIKAIMNLVTGVTSLTFNPTEILAIASEKMKEAVRLVHLPSTVFSNFPVINKKNISHVHTMDFSPRSGYFALGNEKGKALMYRLHHYSDF (SEQ ID NO:739);

and/or KINGRVAASTFSSDSKKVYASSGDGEVYVWDVNSRKCLNRFVDEGSL
YGLSIATSRNGQYVACGSNCGVVNIYNQDSCLEQETNPKPIKAIMNLVTGVTSLT
FNPTTEILAIASEKMKEAVRLVHLPSCVFSNFPVIKNKNISHVHTMDFSPRSG
YFALGNEKGKAL (SEQ ID NO:740).

- 5 This gene is expressed primarily in epididymus and endometrial tumors and to a lesser extent in T cell lymphoma and cell lines derived from colon cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors of the reproductive organs including testis and endometrial cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 411 as residues: Ser-67 to Lys-72, Val-87 to Leu-93, Tyr-128 to Pro-141, Asp-204 to Gly-210.

The tissue distribution indicates that the protein products of this gene are useful for treating tumors of the endometrium or epithelial tumors of the reproductive system.

25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 179**

Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MRILQLILLALATGLVGGETRIIKGFECKLHSQPWQAALFEKTRLLCGATLIAPR
WLLTAAHCLKPRYIVHLGQHNLQKEEGCEQTRTATESFPHPGFNNSLPNKDH
30 RNDIMLVKMASPV SITWAVRPLTLSSRCV TAGTSCSFPAGAAPDP SYACLTPC
DAPTSPSLSTRSVRTP TPATSQTPWCVPACRKGARTPARVTPGALWSVTSLFKA
LSPGARIRVRSPESLVSTRKSANMWTGSRRR (SEQ ID NO:741); ETRIIKGFEC
KLHSQPWQAALFEKTRLLCGATLIAPRWLLTAAHCLKPRYIVHLGQHNLQKEE
GCEQTRTATESFPHPGFNNSLPNKDHRNDIMLVKMASPV SITWAVRPLTLSSR
35 CVTAGTSCSFPAGAAPDP SYACLTPCDAPTSPSLSTRSVRTP TPATSQTPWCVP
ACRKGARTPARVTPGALWSVTSLFKALSPGARIRVRSPESLVSTRKSANMWTG

SRRR (SEQ ID NO:742); or CKLHSQPWQAALFEKTRLLCGATLIAPRWLLT
AAHCLKPRYIVHLGQHNLQKEEGCEQTRTATESFPHPGFNS

(SEQ ID NO:743). The translation product of this gene shares sequence homology
with neuropsin a novel serine protease which is thought to be important in modulating
extracellular signaling pathways in the brain. Owing to the structural similarity to other
serine proteases the protein products of this gene are expected to have serine protease
activity which may be assayed by methods known in the art and described elsewhere
herein.

This gene is expressed primarily in endometrial tumor and to a lesser extent in
colon cancer, benign hypertrophic prostate, and thymus.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, cancers of the endometrium or colon and benign hypertrophy of the
prostate. Similarly, polypeptides and antibodies directed to these polypeptides are
useful in providing immunological probes for differential identification of the tissue(s)
or cell type(s). For a number of disorders of the above tissues or cells, particularly of
the urogenital or reproductive systems, expression of this gene at significantly higher or
lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
another tissue or cell sample taken from an individual having such a disorder, relative to
the standard gene expression level, i.e., the expression level in healthy tissue or bodily
fluid from an individual not having the disorder. Preferred epitopes include those
comprising a sequence shown in SEQ ID NO: 412 as residues: Gly-12 to Ser-22, Pro-
34 to Ser-53.

The tissue distribution and homology to serine proteases indicates that
polynucleotides and polypeptides corresponding to this gene are useful for diagnosing
or treating hyperproliferative disorders such as cancer of the endometrium or colon and
hyperplasia of the prostate.

FEATURES OF PROTEIN ENCODED BY GENE NO: 180

Preferred polypeptide encoded by this gene comprise the following amino acid
sequence: VLQGRYFSPILEMRRLRPEGXXNLPGGSRAQKEPRQDLTLVLWPHC
PHFAMTRSYPVTKQCMVQGSFYCIFKGPVQNWC (SEQ ID NO:744).

Polynucleotides encoding such polypeptide are also provided.

This gene is expressed primarily in fetal brain

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, identifying and expanding stem cells in the CNS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for detecting and expanding stem cell populations in the (or of the) central nervous system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 181

This gene is expressed primarily in early stage human brain and a stromal cell line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities of the CNS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 414 as residues: Gln-42 to Gln-47, Gln-54 to Pro-60.

The tissue distribution indicates that the protein products of this gene play a role in the development of the central nervous system. Therefore this gene and its products

are useful for diagnosing or treating developmental abnormalities of the central nervous system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 182

5 Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MPIIDQVNPELHDFMQSAEVGTIFALSWLITWFGHVLSDFRHVVRLYDF
FLACHPLMPIYFAAVIVLYREQEVLDCDCDMASVHHLLSQIPQDLPYETLISRXE
TFLFSFHPNLLGRPLNSKLRGRQPLLSKTLSTWHQPSRGLIWCCSGXRGLL
10 RPEDRTKDVLTKPRTNRFVKLAVMGLTVALGAAALAVVKSALWAPKFQLQL
FP (SEQ ID NO:745); or CPEFFIPATLPCPFVFAFTSEASSRAYLTQRGPGGLAQ
NLMPLPVGFWMGSLPPPWCWRKWVSEACSCFC (SEQ ID NO:746) These

polypeptides are structurally similar to various TGF-beta family members. Thus, this polypeptide is expected to have a variety of activities in the modulation of cell growth
15 and proliferation.

This gene is expressed primarily in osteoclastoma, microvascular endothelium, and bone marrow derived cell lines.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematological diseases particularly involving aberrant proliferation of stem cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of
25 the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
30 individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 415 as residues: Ser-33 to Ala-39.

The tissue distribution indicates that the protein products of this gene is useful for treating disorders of the progenitors of the immune system. Applications include in vivo expansion of progenitor cells, ex vivo expansion of progenitor cells, or the
35 treatment of tumors of the circulatory system, such as lymphomas.

FEATURES OF PROTEIN ENCODED BY GENE NO: 183

This gene maps to chromosome 17 and therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 17. In specific embodiments, polypeptides of the invention comprise the sequence:

- 5 GFGSVSAAGRRSGGTWQPVQ (SEQ ID NO:747); PGGLAVGSRWWSRSLT (SEQ ID NO:748); LEPSRQRRPRRRGGTSRPETDQRAKCWRQL (SEQ ID NO:749); and/or VCLRCQNRMEN (SEQ ID NO:750). In further specific embodiments, polypeptides of the invention comprise the sequence: MAACTARRPGR GQPLVVPVADXGPVAKAALCAAXAGAFSPASTTTTRRHLSSNRNPEGKVLETV
- 10 GVFEVPKQNGKYETGQLFLHSIFGYRGVVLFPWQARLXDRDVASAAPEKAEN PAGHGSKEVKGKTHTYQVLIDARDCPHISQRSQTEAVTFLANHDDSRALYAIP GLDYVSHEDILPYTSTDQVPIQHELPERFLLYDQTKAPPFVARETLRAWQEKNH PWLELSDVHRETTENIRVTVIPFYMGMREAQNSHVYWWRYCIRLENLSDSDVVQ LRERHWRIFSLSGTLETVRGRGVVGREPVLSKEQPAFQYSSHVSLQASSGHMW
- 15 GTFRFERPDGSHFDVRIPPSLESNKDEKTPPSGLHW (SEQ ID NO:751); MAACTARRPGRGQPLVVPVADXGPVAKAALCAA (SEQ ID NO:752); VLETVG VFEVPKQNGKYETGQLFLHSIFGYRGVVL (SEQ ID NO:757); GLDYVSHEDILPYTST (SEQ ID NO:758); DVHRETTENIRVTVIPFYM (SEQ ID NO:759); WWRYCIRLENLSDSDVQLRER (SEQ ID NO:760); and/or PAFQYSS
- 20 HVSLQASSGHMWGTFRFER (SEQ ID NO:761). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in gall bladder, prostate, and fetal brain, and to a lesser extent in a few tumor and fetal tissues.

- Therefore, polynucleotides and polypeptides of the invention are useful as
- 25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, growth related disorders such as cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders
- 30 of the above tissues or cells, particularly of the prostate, gall bladder, and fetal brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,
- 35 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of growth-related disorders, such cancers.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 184

In specific embodiments, polypeptides of the invention comprise the sequence:SLCCPEGAEGC (SEQ ID NO:762) and/or QLKKTHYDRPCP (SEQ ID NO:763). Polynucleotides encoding these polypeptides are also encompassed by the invention.

10 This gene is expressed primarily in stromal cell, tonsil, and glioblastoma and to a lesser extent in some other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and inflammatory disorders and glioblastoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the stromal cells, tonsil, and glioblastoma expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Additionally, it is believed that the product of this gene regulates pancreatic cell differentiation into beta cells. Accordingly, polynucleotides and polypeptides of the invention are useful in the treatment of insulin-dependent diabetes mellitus and associated conditions e.g. pancreatic hypofunction and the prevention, as well as the treatment of undifferentiated type pancreatic cancers. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 417 as residues: Pro-27 to Ala-32.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune and inflammatory disorders and glioblastoma.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 185

This gene is expressed primarily in hepatocellular carcinoma and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, liver diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the liver, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 418 as residues: Gly-32 to Lys-39.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of liver diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 186

This gene is expressed primarily in hippocampus and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hippocampus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 187

This gene is expressed primarily in bone cancer and hippocampus and to a lesser extent in osteoclastoma and other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, bone-related disorders and neuronal diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, osteoclast, and hippocampus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of bone-related disorders and neuronal diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 188

This gene maps to chromosome 4 and therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 4.

This gene is expressed primarily in neuronal tissues such as hippocampus, spinal cord, and hypothalamus and to a lesser extent in a few other tissues such as ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 189

This gene maps to chromosome 10, therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 10.

5 This gene is expressed primarily in neuronal tissues and immune tissues, and to a lesser extent in a few other tissues such as skin tumor, lung etc.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal and immune-related disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal and immune-related tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 422 as residues: Pro-19 to Asp-25.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal and immune-related disorders.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 190

The translation product of this gene shares sequence homology with human N33, a gene located in a homozygously deleted region of human metastatic prostate cancer which is thought to be important in prevention of prostate cancer. In specific embodiments, polypeptides of the invention comprise the sequence:

30 AQRKKEMVLSEKVSQLEWTKRVPVIRMGDKFRRLVKAPPRNYSVIVMFTA
LQLHRQCVVCKQADEEFQILANSWRYSSAFTNRIFFAMVDFDEGSDVFQMLNM
NSAPTFINFPKKGPKRGDTYELQVRGFSAEQIARWIADRTDVNIRVIRPPNMA
ARWRFWCVSVT (SEQ ID NO:765); MVVALLIVCDVPSAS (SEQ ID NO:766);
AQRKKEMVLSEKVSQLEWTKRVPVIRMGDKF (SEQ
35 ID:768); RRLVKAPPRNYSVIVMFTALQLHRQCVVCKQADEEFQILANSWRY
SSAFTNRIFFA (SEQ ID NO:769); MVDFDEGSDVFQMLNMNSAPTFINFPK
GKP (SEQ ID NO:770); KRGDTYELQVRGFSAEQIARWIADRTDVNIRVIRPPN

(SEQ ID NO:771); and/or YAGPLMLGLLLA VIGGLVYLRRVTWNFSLIKLDGLLQL CVLCLL (SEQ ID NO:772). Polynucleotides encoding these polypeptides are also encompassed by the invention.

5 This gene is expressed primarily in infant adrenal gland prostate cell line and to a lesser extent in a few other tissues like liver, smooth muscle etc.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, prostate cancer and endocrine disorders. Similarly, polypeptides and
10 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate and adrenal gland, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
15 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 423 as residues: Pro-34 to Gly-43, Arg-113 to Pro-120.

20 The tissue distribution and homology to N33 indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for prostate cancer and endocrine disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 191

25 This gene is expressed primarily in T cell and to a lesser extent in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to
30 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal
35 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue

or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 424 as residues: Trp-3 to Phe-9.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 192

This gene maps to chromosome 6, therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 6. Neural activity and neurotrophins induce synaptic remodeling in part by altering gene expression. This gene is believed to be a glycosylphosphatidylinositol-anchored protein encoded by a hippocampal gene and to possess neural activity. This molecule is believed to be expressed in postmitotic-differentiating neurons of the developing nervous system and neuronal structures associated with plasticity in the adult. Message of this gene is believed to be induced by neuronal activity and by the activity-regulated neurotrophins BDNF and NT-3. The product of this gene is believed to stimulate neurite outgrowth and arborization in primary embryonic hippocampal and cortical cultures and to act as a downstream effector of activity-induced neurite outgrowth. In specific embodiments, polypeptides of the invention comprise the sequence: DAVFKGFSDCLLKLGD (SEQ ID NO:773); CQEGAKDMWDKLRKESKNLN (SEQ ID NO:774); VLLVSLSAALATWLSF (SEQ ID NO:775); MGLKLNGRYISLILAVQIAYLVQAVR AAGKCDVFKGFSDCLLKLGD (SEQ ID NO:776); PAAWDDKTNIKTVC TYW EDFHSCTVTALTDCQEGAKDMWDKLRKESKNLN IQGSLFELCGSGNGAAGSL LPAFPVLLVSLSAALATWLSF (SEQ ID NO:777); and/or MGLKLNGRYISLILAVQIAYLVQAVRAAGKCDVFKGFSDCLLKLGD SXXXXXPAAWDDKTNIKTVC TYW EDFHSCTVTALTDCQEGAKDMWDKLRKESKNLN IQGSLFELCGSGNGAAGSL LPAFPVLLVSLSAALATWLSF (SEQ ID NO:778). Polynucleotides encoding this polypeptide are also encompassed by the invention.

This gene is expressed primarily in human placenta, endometrial tumor and tissues of the central nervous system (CNS).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, relating to reproductive disorders, cancers and neurological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and neurological disorders, expression of this gene at significantly higher

or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 425 as residues: Asp-47 to Asp-63, His-75 to Tyr-80, Pro-83 to Tyr-89.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of reproductive disorders such as endometrial tumors. Expression of this gene in tissues of the CNS and its strong homology to Neuritin suggest that the protein product from this gene may also be used in the treatment and diagnosis of neurological disorders and in the regeneration of neural tissues, e.g., following injury.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 193**

The translation product of this gene shares sequence homology with tenascin which is thought to be important in development. The translation product of this gene is believed to be a ligand of the fibroblast growth factor family. FGF ligand activity is known in the art and can be assayed by methods known in the art and disclosed elsewhere herein.

This gene is expressed primarily in endometrial tumors, and other types of tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 426 as residues: Gly-29 to Glu-34, Arg-71 to Arg-76, Thr-176 to Cys-182, Gly-184 to Glu-199.

The tissue distribution and homology to tenascin indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers.

5 **FEATURES OF PROTEIN ENCODED BY GENE NO: 194**

In specific embodiments, polypeptides of the invention comprise the sequence: MNSAAGFSLDRRERVLKLGESFEKQPRCASTLC (SEQ ID NO:779).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in fetal human lung and neutrophils.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, lung development and respiratory disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes
15 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the respiratory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual
20 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

 The tissue distribution in fetal lung and neutrophils indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis
25 and treatment of lung and immunity related diseases, for example, lung cancer, viral, fungal or bacterial infections (e.g. lesions caused by tuberculosis), inflammation (e.g. pneumonia), metabolic lesions etc.

FEATURES OF PROTEIN ENCODED BY GENE NO: 195

30 This gene is expressed primarily in breast lymph node.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunal disorders. Similarly, polypeptides and antibodies directed to
35 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immunal disorders.

10 **FEATURES OF PROTEIN ENCODED BY GENE NO: 196**

This gene maps to chromosome 5 and accordingly, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 5. The translation product of this gene shares sequence homology with human M-phase phosphoprotein 4 which is thought to be important in phosphorylation and signal transduction processes. In specific embodiments, polypeptides of the invention comprise the sequence:

15 TIYPTEELQAVQKIVSITERALKLVSD (SEQ ID NO:780); RALKGVLRV
GVLAKGLLLRGDRNVNLVLLC (SEQ ID NO:781); ALAALRHAKWFQARAN
GLQSCVIRILRDLQCQRVPTWS (SEQ ID NO:782); GDALRRVFECISSGIL (SEQ
ID NO:783); LAFRQIHKVLGMDPLP (SEQ ID NO:784); and/or TIYPTEELQAVQ
20 KIVSITERALKLVSDSLSEHEKNKNKEGDDKKEGGKDRALKGVLRVGVLAGK
LLLRGDRNVNLVLLCSEKPSKTLLSRIAENLPKQLAVISPEKYDIKCAVSEAAIIL
NSCVEPKMQVTITLTSPIREENMREGDVTSGMVKDPPDVLDRQKCLDALAALR
HAKWFQARANGLQSCVIRILRDLQCQRVPTWSDFPSWAMELLVEKAISSASSP
QSPGDALRRVFECISSGILKGSPGLLDPCCKDPFDTLATMTDQQREDITSSAQFA
25 LRLLAFRQIHKVLGMDPLPQMSQRFNIHNNRKRRRRSDGVDGFEGGKDKK
DYDNF (SEQ ID NO:785). Polynucleotides encoding these polypeptides are also
encompassed by the invention.

This gene is expressed primarily in Human Hippocampus and to a lesser extent in Prostate, Human Frontal Cortex.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders related to reproductive system and nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
35 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system and nervous system, expression of this gene at significantly higher or lower

- levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to human M-phase phosphoprotein 4 indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of reproductive and nervous system disorders.

10 **FEATURES OF PROTEIN ENCODED BY GENE NO: 197**

In specific embodiments, polypeptides of the invention comprise the sequence: MGSQHSAAARPSSCRKQEDDRDG (SEQ ID NO:786); LLAEREQEEAIAQFPYVEFTGRDSITCLTC (SEQ ID NO:787); and/or QGTGYIPTEQVNELVALIPHSQRLRPQRTKQYV (SEQ ID NO:788).

- 15 Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in Human Primary Breast Cancer and to a lesser extent in Human Adult Spleen, Hodgkin's Lymphoma I, Salivary Gland.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and immunal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 430 as residues: Ser-126 to Gly-138.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and immunal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 198

This gene is expressed primarily in monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, blood cell disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of blood cell disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 199

This gene is expressed primarily in Human Ovary and Synovia and to a lesser extent in Human 8 Week Whole Embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, reproductive and developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and developmental system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of reproductive and developmental disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 200

This gene maps to chromosome 8 and therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 8. The translation product of this gene shares limited sequence homology with collagen proline rich domain.

This gene is expressed primarily in CNS.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 433 as residues: Pro-35 to Asp-41.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 201

Translation product of this gene shares homology with a mammalian histone H1a protein. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: ARLNVGRESLKREMLKSQGVKVSSEPMGAR HSSWPEGAAFCCKKVQGAQMFPFRR (SEQ ID NO:789); ARLNVGRESLKR EML (SEQ ID NO:790); LKSQGVKVSSEPMGARHSSW (SEQ ID NO:791); AFCCKKVQGAQMFPFRR (SEQ ID NO:792). An additional embodiment is the polynucleotide fragments encoding these polypeptide (See Accession No. pirlS24178) fragments.

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at

5 significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in vital immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such

15 as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 202

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as

20 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above

25 tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level

30 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for

35 immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 203

This gene is expressed primarily in Neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, infectious disorders, immune disorders, and cancers. Similarly,
polypeptides and antibodies directed to these polypeptides are useful in providing
immunological probes for differential identification of the tissue(s) or cell type(s). For
10 a number of disorders of the above tissues or cells, particularly of the immune system,
expression of this gene at significantly higher or lower levels may be routinely detected
in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
an individual having such a disorder, relative to the standard gene expression level, i.e.,
15 the expression level in healthy tissue or bodily fluid from an individual not having the
disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
NO: 436 as residues: Thr-31 to Lys-36.

The tissue distribution indicates that polynucleotides and polypeptides
corresponding to this gene are useful for diagnosis and treatment of infectious
20 disorders, immune disorders, and cancers. Since the gene is expressed in cells of
lymphoid origin, the natural gene product may be involved in immune functions.
Therefore it may be also used as an agent for immunological disorders including
arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. Protein, as
well as, antibodies directed against the protein may show utility as a tumor marker
25 and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 204

This gene maps to chromosome 16 and therefore polynucleotides of the
invention can be used in linkage analysis as markers for chromosome 16. The
30 translation product of this gene shares sequence homology with lactate dehydrogenase
which is thought to be important in lactate metabolism.

This gene is expressed primarily in human tonsils and to a lesser extent in
Spleen, and Neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
35 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, immune disorders, infectious disorders, and cancers. Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune disorders, infectious disorders, and cancers, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 437 as residues: Gly-7 to Ser-12.

The tissue distribution and homology to lactate dehydrogenase gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders, infectious disorders, and cancers.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 205**

The translation product of this gene shares sequence homology with Gcap1 protein which is developmentally regulated in brain.

This gene is expressed primarily in placenta and endometrial tumor and to a lesser extent in several other tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, vasculogenesis/angiogenesis and tumorigenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system and tumors, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to Gcap1 protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorder or dysfunction of vascular system of tumorigenesis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 206

In specific embodiments, polypeptides of the invention comprise the sequence
 MPYAQWLAENDRFEEAQKAFHKAGRQREA (SEQ ID NO:799);
 VQVLEQLTNNAVAESRFNDAAYYYWMLSMQCLDIAQD (SEQ ID NO:794);
 5 PAQKDTMLGKFYHFQRLAELYHGYHAIHRHTEDP (SEQ ID NO: 795);
 FSVHRPETLFNISRFLHSLPKDTPSGISKVKILFT (SEQ ID NO:800);
 LAKQSKALGAYRLARHAYDKLRGLYIP (SEQ ID NO:796); ARFQKSIELG
 TLTIRAKPFHDSEELVPLCYRCSTNN (SEQ ID NO: 797); and/or PLLNNLGNVC
 INCRQPFISSASSYDVLHLVEFYLEEGITDEEAISLIDLEVLRPKRDDRQLEICKQQ
 10 LPDSCG (SEQ ID NO:798). Polynucleotides encoding these polypeptides are also
 encompassed by the invention.

This gene is expressed primarily in testes.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 15 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, male reproductive and endocrine disorders. Similarly, polypeptides and
 antibodies directed to these polypeptides are useful in providing immunological probes
 for differential identification of the tissue(s) or cell type(s). For a number of disorders
 of the above tissues or cells, particularly of the reproductive and endocrine systems,
 20 expression of this gene at significantly higher or lower levels may be routinely detected
 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
 an individual having such a disorder, relative to the standard gene expression level, i.e.,
 the expression level in healthy tissue or bodily fluid from an individual not having the
 25 disorder.

The tissue distribution indicates that polynucleotides and polypeptides
 corresponding to this gene are useful for treatment of male reproductive and endocrine
 disorders.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 207

This gene is expressed in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 35 not limited to, lung diseases such as cystic fibrosis. Similarly, polypeptides and
 antibodies directed to these polypeptides are useful in providing immunological probes
 for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the respiratory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual
5 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 440 as residues: Tyr-49 to Cys-54.

The tissue distribution indicates that polynucleotides and polypeptides
10 corresponding to this gene are useful for detection and treatment of disorders associated with developing lungs particularly in premature infants where the lungs are the last tissues to develop. The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of lung tumors since the gene may be involved in the regulation of cell division,
15 particularly since it is expressed in fetal tissue. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
1	HLHDS67	97979 03/27/97	Uni-ZAP XR	11	2526	427	2526	458	458	234	1	30	31	30
2	HLHDZ58	97979 03/27/97	Uni-ZAP XR	12	1131	1	1131	129	129	235	1	14	15	115
3	HLMMJ13	97979 03/27/97	Lambda ZAP II	13	941	39	941	62	62	236	1	44	45	102
3	HLMMJ13	97979 03/27/97	Lambda ZAP II	218	941	39	941	245	245	441	1	35	36	41
4	HLTEI25	97979 03/27/97	Uni-ZAP XR	14	843	1	843	155	155	237	1	19	20	42
5	HMSJX24	97979 03/27/97	Uni-ZAP XR	15	1018	1	1018	90	90	238	1	18	19	36
6	HNFEID65	97979 03/27/97	Uni-ZAP XR	16	661	1	661	76	76	239	1	28	29	127
7	HNHDX07	97979 03/27/97	Uni-ZAP XR	17	553	1	553	106	106	240	1	23	24	66
8	HNHGC82	97979 03/27/97	Uni-ZAP XR	18	869	1	869	101	101	241	1	21	22	68
9	HNHGO09	97979 03/27/97	Uni-ZAP XR	19	959	1	959	176	176	242	1	21	22	44
10	HOUBE18	97979 03/27/97	Uni-ZAP XR	20	1446	1	1446	101	101	243	1	27	28	50
11	HOUDL69	97979 03/27/97	Uni-ZAP XR	21	1471	579	1460	692	692	244	1	31	32	42

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
12	HPMF171	97979 03/27/97	Uni-ZAP XR	22	1402	242	1402	401	401	245	1	32	33	60
13	HPMGQ55	97979 03/27/97	Uni-ZAP XR	23	1047	1	1047	164	164	246	1	26	27	35
14	HPQAC69	97979 03/27/97	Lambda ZAP II	24	990	1	988	82	82	247	1	20	21	37
15	HPTBB03	97979 03/27/97	Uni-ZAP XR	25	1208	350	1173	398	398	248	1	29	30	210
16	HPTWA66	97979 03/27/97	pBluescript	26	1922	1381	1922	24	24	249	1	33	34	547
16	HPTWA66	97979 03/27/97	pBluescript	219	575	1	575	148	148	442	1	22	23	65
17	HPTWC08	97979 03/27/97	pBluescript	27	1951	1422	1874	219	219	250	1	19	20	299
18	HRGCZ46	97979 03/27/97	Uni-ZAP XR	28	3989	2635	3989		2748	251	1	16	17	39
19	HSADV34	97979 03/27/97	Uni-ZAP XR	29	3735	2966	3735	272	272	252	1	30	31	594
19	HSADV34	97979 03/27/97	Uni-ZAP XR	220	3018	1929	3018	26	26	443	1	1	2	156
20	HSDFW61	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	30	1667	59	1625	138	138	253	1	32	33	130
21	HSDGP60	97974 04/04/97	Uni-ZAP XR	31	1408	1	1408	285	285	254	1			20

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of NT Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
22	HSOAJ55	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	32	2031	1273	2031	1285	1285	255	1	29	30	30
23	HSQEO84	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	33	971	13	971	91	91	256	1	19	20	218
23	HSQEO84	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	221	968	8	968	86	86	444	1	20	21	56
24	HSXAM05	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	34	1792	369	1792	470	470	257	1	26	27	49
25	HSXAS67	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	35	896	1	896	96	96	258	1	32	33	121
26	HTDAF28	97974 04/04/97 209080 05/29/97	pSport1	36	912	1	912	38	38	259	1	22	23	87
27	HTEGO64	97974	Uni-ZAP XR	37	1382	67	1382	271	271	260	1			25

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		04/04/97 209080 05/29/97												
28	HTGEU09	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	38	872	1	872	74	74	261	1	18	19	28
29	HTOAM21	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	39	812	1	812	41	41	262	1	30	31	43
30	HTPBW79	209511 12/03/97	Uni-ZAP XR	40	1515	118	1507	302	302	263	1	24	25	362
30	HTSEV09	97974 04/04/97 209080 05/29/97	pBluescript	222	1404	1	1265	92	92	445	1	19	20	415
31	HJPCD40	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	41	704	22	704		117	264	1	18	19	127
32	HTWB48	97974 04/04/97 209080 05/29/97	pSport1	42	1094	1	1094	32	32	265	1	34	35	53
33	HTWC146	97974 04/04/97	pSport1	43	1821	892	1647	56	56	266	1	26	27	28

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
34	HTXG175	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	44	1024	30	1024		167	267	1	20	21	25
35	HWTFB59	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	45	983	779	983	85	85	268	1	30	31	221
35	HWTFB59	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	223	707	488	707	514	514	446	1	41	42	64
36	HADA74	97974 04/04/97 209080 05/29/97	pSport1	46	2421	664	1587	710	710	269	1			2
37	HAGFB60	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	47	840	1	840	97	97	270	1	30	31	48
38	HATEF60	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	48	2432	1193	2246	1491	1491	271	1	17	18	51
39	HBMSN25	97974 05/29/97	Uni-ZAP XR	49	1742	1165	1742	1207	1207	272	1	23	24	31

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
40	HCDAR68	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	50	1487	181	1455	325	325	273	1	35	36	56
41	HCE3179	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	51	1328	251	1328	525	525	274	1			21
42	HMDAN54	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	52	1856	725	1853	928	928	275	1	33	34	50
43	HCECA49	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	53	1558	310	1408	393	393	276	1			1
44	HCEEC15	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	54	948	1	948	9	9	277	1	23	24	65
45	HCESEF40	97974 04/04/97 209080 05/29/97	pBluescript	55	990	99	990	193	193	278	1	32	33	256

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
45	HCESF40	97974 04/04/97 209080 05/29/97	pBluescript	224	1384	99	1384	193	193	447	1	32	33	205
46	HCFMV39	97974 04/04/97 209080 05/29/97	pSport1	56	1603	1	1296	96	96	279	1	29	30	102
47	HCMsX86	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	57	1052	5	786	12	12	280	1	28	29	32
48	HcNAP62	97975 04/04/97 209081 05/29/97	Lambda ZAP II	58	814	1	558	93	93	281	1	22	23	42
49	HCRAF32	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	59	1215	257	1215		356	282	1	19	20	20
50	HCUDC07	97975 04/04/97 209081 05/29/97	ZAP Express	60	478	1	478	147	147	283	1	36	37	69
51	HCWBB42	97975 04/04/97 209081	ZAP Express	61	618	1	618	212	212	284	1	35	36	74

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
52	HDTAB05	97975 04/04/97 209081 05/29/97	pcMVSPORT 2.0	62	751	1	751	257	257	285	1	21	22	32
53	HE2AV74	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	63	780	283	780		433	286	1			16
54	HE2AV71	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	64	588	21	588	169	169	287	1			16
55	HE2GS36	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	65	774	272	774	445	445	288	1			37
56	HE2OF09	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	66	1866	1313	1866	1596	1596	289	1			11
57	HE6EU50	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	67	1152	117	686	237	237	290	1	20	21	34
58	HE9HU17	97975 04/04/97	Uni-ZAP XR	68	2483	1577	2448	1620	1620	291	1			14

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
59	HE9ND48	97975 04/04/97 209081	Uni-ZAP XR	69	536	1	536	83	83	292	1	36	37	43
60	HEBBW11	97975 04/04/97 209081	Uni-ZAP XR	70	865	647	865		388	293	1	30	31	135
61	HELDY74	97975 04/04/97 209081	Uni-ZAP XR	71	932	1	932	201	201	294	1	17	18	33
62	HEMAE80	97975 04/04/97 209081	Uni-ZAP XR	72	996	1	945	12	12	295	1	24	25	136
63	HFEBA88	97975 04/04/97 209081	Uni-ZAP XR	73	785	464	785	356	356	296	1	29	30	57
64	HFGAB89	97975 04/04/97 209081	Uni-ZAP XR	74	1069	196	1047	295	295	297	1	32	33	34
65	HFVHY45	97975	pBluescript	75	831	1	831		89	298	1	30	31	76

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
66	HGBAJ93	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	76	590	1	590	233	233	299	1	38	39	94
67	HGBBQ69	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	77	1274	1	1273	105	105	300	1	24	25	43
68	HHFCF08	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	78	1133	4	1042	175	175	301	1	23	24	30
69	HHFHJ59	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	79	661	1	661	192	192	302	1	29	30	112
70	HHFHR32	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	80	1378	1	1378		358	303	1			13
71	HHGCN69	97975 04/04/97 209081 05/29/97	Lambda ZAP II	81	1440	298	1440	532	532	304	1	23	24	34

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
72	HHGDO13	97975 04/04/97 209081 05/29/97	Lambda ZAP II	82	1381	766	1371	993	993	305	1	23	24	34
73	HHPEFD63	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	83	1706	182	1644	257	257	306	1	24	25	81
74	HHSEG23	97976 04/04/97	Uni-ZAP XR	84	573	1	573	160	160	307	1	18	19	71
75	HJPAV06	97976 04/04/97	Uni-ZAP XR	85	684	199	684	323	323	308	1	27	28	33
76	HKIXL73	97976 04/04/97	pBluescript	86	1036	591	1036	690	690	309	1	32	33	114
77	HKMNC43	97976 04/04/97	pBluescript	87	908	1	908	139	139	310	1	18	19	108
78	HMEJE31	97976 04/04/97	Lambda ZAP II	88	655	1	655	165	165	311	1	33	34	64
79	HMSKS35	97976 04/04/97	Uni-ZAP XR	89	1102	1	1102	228	228	312	1	26	27	49
80	HNFAE54	97976 04/04/97	Uni-ZAP XR	90	1533	665	1518	347	347	313	1	26	27	293
81	HNFJH45	97976 04/04/97	Uni-ZAP XR	91	575	1	575	275	275	314	1	30	31	67
82	HNGBT31	97976 04/04/97	Uni-ZAP XR	92	639	1	639	224	224	315	1	28	29	104

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
83	HNGIN60	97976 04/04/97	Uni-ZAP XR	93	744	1	744	225	225	316	1	43	44	70
84	HNGJG84	97976 04/04/97	Uni-ZAP XR	94	526	1	526	268	268	317	1	29	30	38
85	HNHDW42	97976 04/04/97	Uni-ZAP XR	95	426	1	426	168	168	318	1	28	29	71
86	HNHFL57	97976 04/04/97	Uni-ZAP XR	96	844	1	844	98	98	319	1	25	26	61
87	HOGAR52	97977 04/04/97 209082 05/29/97	PCMVSPORT 2.0	97	1985	453	1985	533	533	320	1	17	18	285
88	HOSBZ55	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	98	1416	69	1416	246	246	321	1	32	33	54
89	HOSDI92	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	99	1935	141	772		274	322	1	20	21	58
90	HPBCU51	97977 04/04/97 209082 05/29/97	pBluescript SK-	100	599	1	599	86	86	323	1	27	28	119
91	HPCAL49	97977 04/04/97 209082	Uni-ZAP XR	101	784	1	784		280	324	1	18	19	43

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
92	HPFCR13	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	102	1035	602	1035	859	859	325	1	32	33	58
93	HPHAC83	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	103	2218	840	2182	1035	1035	326	1	17	18	17
94	HPMBQ32	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	104	1351	1	1351	18	18	327	1	23	24	86
95	HPWAN23	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	105	2066	51	2052	270	270	328	1	29	30	537
95	HPWAN23	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	226	2057	1	1954	220	220	449	1	29	30	315
96	HRDFB85	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	106	1705	23	1697	233	233	329	1	21	22	201
97	HRGBR28	97977 04/04/97	Uni-ZAP XR	107	1167	611	1167	53	53	330	1	1	2	263

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		209082 05/29/97												
98	HSKGN81	97977 04/04/97 209082 05/29/97	pBluescript	108	1907	151	1432	353	353	331	1	23	24	260
98	HSKGN81	97977 04/04/97 209082 05/29/97	pBluescript	227	2084	335	2084	537	537	450	1	19	20	23
99	HSPAHS6	97977 04/04/97 209082 05/29/97	pSport1	109	611	1	576	229	229	332	1	25	26	47
100	HE8EU04	209746 04/07/98	Uni-ZAP XR	110	2632	294	2632	337	337	333	1	25	26	333
100	HSXBT86	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	228	2143	53	1096	235	235	451	1			9
101	HSXCS62	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	111	2249	1	1953	90	90	334	1	18	19	199
102	HTEFU09	97977 04/04/97 209082	Uni-ZAP XR	112	2198	228	2158	400	400	335	1			23

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of First Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF	
103	HTBKM35	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	113	1043	40	1043	320	320	336	1	20	21	142
104	HTGEP89	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	114	703	1	703	285	285	337	1	29	30	94
105	HTGEW91	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	115	3684	526	1338	584	584	338	1	24	25	37
106	HTOEY16	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	116	1965	127	1915	202	202	339	1	27	28	38
107	HTPCN79	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	117	503	1	503		1	340	1	7	8	70
108	HTSGM54	97977 04/04/97 209082 05/29/97	pBluescript	118	1133	316	1069		423	341	1	12	13	84
109	HTSHE40	97977 04/04/97	pBluescript	119	1101	118	956	218	218	342	1	31	32	89

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
110	HTWAF58	97977 04/04/97 209082 05/29/97	Lambda ZAP II	120	282	1	282	137	137	343	1	25	26	48
111	HTWBY29	97977 04/04/97 209082 05/29/97	pSport1	121	2635	1593	2489	1654	1654	344	1	25	26	55
112	HUKFC71	209007 04/28/97 209083 05/29/97	Lambda ZAP II	122	994	1	932		272	345	1	15	16	221
113	HCE3Q10	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	123	1542	1	1542	143	143	346	1	25	26	63
114	HCEVR60	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	124	1390	82	1390	127	127	347	1	32	33	153
115	HDTAW95	209007 04/28/97 209083 05/29/97	pcMVSPORT 2.0	125	1288	412	1288	571	571	348	1			16
116	HE6EL90	209007	Uni-ZAP XR	126	1517	1	1452	243	243	349	1			9

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
117	HELB29	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	127	1073	198	1073		776	350	1			13
118	HERAH36	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	128	300	155	300	202	202	351	1			17
119	HFXBW82	209007 04/28/97 209083 05/29/97	Lambda ZAP II	129	1275	1	1275	56	56	352	1	23	24	61
120	HHP1D20	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	130	472	51	472		243	353	1			32
121	HIBED17	209007 04/28/97 209083 05/29/97	Other	131	1950	284	1927	395	395	354	1	72	73	245
122	HLTER03	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	132	990	1	990	78	78	355	1	22	23	34

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
123	HOABL56	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	133	1720	565	1720	660	660	356	1	18	19	21
124	HPMCJ92	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	134	705	28	705	106	106	357	1	28	29	98
125	HPWAZ95	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	135	323	1	323	88	88	358	1	27	28	78
126	HRGBR18	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	136	582	1	582		16	359	1	17	18	30
127	HSUBW09	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	137	1021	1	1021	153	153	360	1	32	33	56
128	HUKCO64	209007 04/28/97 209083 05/29/97	Lambda ZAP II	138	1777	439	1777		521	361	1			2
129	H6EAA53	209007 04/28/97 209083	Uni-ZAP XR	139	643	303	643		313	362	1	7	8	31

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
130	HAGAI11	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	140	1220	1	1220		127	363	1	16	17	27
131	HAGAO39	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	141	721	1	721		415	364	1			14
132	HALSK07	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	142	1468	125	1468	210	210	365	1	29	30	33
133	HALSQ59	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	143	300	4	300	101	101	366	1	22	23	66
134	HAIBP89	unknown 05/18/98	Uni-ZAP XR	144	2243	173	2243	311	311	367	1	27	28	317
134	HBGCB91	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	229	1025	409	1025	624	624	452	1	20	21	25
135	HBMTD81	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	145	1082	163	1082	357	357	368	1			30

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
136	HBXGK12	209008 04/28/97 209084 05/29/97	ZAP Express	146	4313	1153	4313	1313	1313	369	1	18	19	42
137	HFKFI07	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	147	1183	1	1183	149	149	370	1	41	42	254
138	HCOAI40	209008 04/28/97 209084 05/29/97	Lambda ZAP II	148	734	1	734	285	285	371	1			19
139	HCWHZ24	209008 04/28/97 209084 05/29/97	ZAP Express	149	1405	1	1405	108	108	372	1	34	35	63
140	HE2GT20	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	150	2890	1178	2890	1178	1178	373	1	31	32	39
141	HE8EY43	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	151	2399	1181	2399	1265	1265	374	1	30	31	34
142	HFCEB37	209008 04/28/97 209084	Uni-ZAP XR	152	802	352	802		487	375	1			10

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
143	HFTCT67	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	153	461	24	461	145	145	376	1	37	38	63
144	HGLAM46	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	154	2388	818	2388	648	648	377	1			18
145	HHGBR15	209008 04/28/97 209084 05/29/97	Lambda ZAP II	155	642	322	642	400	400	378	1			4
146	HJAU36	209008 04/28/97 209084 05/29/97	pBluescript SK-	156	1251	583	1251		933	379	1	16	17	16
147	HUSIT49	209008 04/28/97 209084 05/29/97	pSport1	157	2127	247	2127	383	383	380	1	47	48	83
148	HKLAB16	209008 04/28/97 209084 05/29/97	Lambda ZAP II	158	1625	817	1625	1012	1012	381	1	18	19	20
149	HLMMU76	209008 04/28/97	Lambda ZAP II	159	1687	1307	1687	1296	1296	382	1	28	29	28

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
150	HMSKQ35	209008 04/28/97	Uni-ZAP XR	160	1842	172	1463	319	319	383	1	30	31	33
		209084 05/29/97												
151	HNHED86	209008 04/28/97	Uni-ZAP XR	161	770	1	770	30	30	384	1	31	32	46
		209084 05/29/97												
152	HNHEJ88	209008 04/28/97	Uni-ZAP XR	162	519	1	519	242	242	385	1	17	18	24
		209084 05/29/97												
153	HNHFQ63	209008 04/28/97	Uni-ZAP XR	163	753	1	753	164	164	386	1	17	18	67
		209084 05/29/97												
154	HOECU83	209009 04/28/97	Uni-ZAP XR	164	1400	189	1400		508	387	1	22	23	33
155	HPTRC15	209009 04/28/97	pBluescript	165	2153	594	2153		611	388	1			13
156	HSKCP69	209009 04/28/97	Uni-ZAP XR	166	1251	219	1120			389	1			
156	HSKCP69	209009 04/28/97	Uni-ZAP XR	230	1250	223	1250	393	393	453	1	32	33	171
157	H6EAE26	209009	Uni-ZAP XR	167	882	48	882	155	155	390	1	33	34	153

Gene No.	cDNA Clone ID	ATCC Deposit Nr. and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
158	HAGBX03	209009 04/28/97	Uni-ZAP XR	168	1208	1	1208	182	182	391	1			8
159	HAGDQ47	209009 04/28/97	Uni-ZAP XR	169	1307	1	1307	44	44	392	1	22	23	60
160	HAICP19	209009 04/28/97	Uni-ZAP XR	170	1624	89	1483	128	128	393	1	18	19	446
161	HAUAE83	209009 04/28/97	Uni-ZAP XR	171	2003	889	2003	1080	1080	394	1			23
162	HBHAD12	209009 04/28/97	Uni-ZAP XR	172	786	1	786		176	395	1	17	18	23
163	HBMTY28	209009 04/28/97	Uni-ZAP XR	173	1758	962	1758	1184	1184	396	1	27	28	34
164	HBMVP04	209009 04/28/97	Uni-ZAP XR	174	888	330	862		546	397	1			2
165	HCDDDB78	209009 04/28/97	Uni-ZAP XR	175	2379	750	2379	901	901	398	1	18	19	24
166	HCEQA68	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	176	1348	1	1348	12	12	399	1	28	29	78
167	HCEZS40	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	177	1502	178	1502	315	315	400	1			20
168	HCFNF11	209010	pSport1	178	1637	26	1607	152	152	401	1	44	45	257

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		04/28/97 209085 05/29/97												
169	HCRBL20	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	179	2911	1103	2858	192	192	402	1	32	33	424
169	HCRBL20	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	231	1811	20	1811	93	93	454	1	36	37	95
170	HCUBL62	209010 04/28/97 209085 05/29/97	ZAP Express	180	519	1	519	57	57	403	1	28	29	32
171	HDSAP81	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	181	968	320	968	476	476	404	1	27	28	79
172	HE2CT29	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	182	1128	1	1128	111	111	405	1	26	27	94
173	HE8MG65	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	183	2276	48	2276	88	88	406	1	37	38	257

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
173	HE8MG65	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	232	2271	56	2232	79	79	455	1	43	44	170
174	HE9FB42	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	184	2500	76	1693	518	518	407	1	1	2	623
175	HEMAM41	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	185	1337	60	1328	175	175	408	1	39	40	190
175	HEMAM41	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	233	1338	33	1327	175	175	456	1	32	33	91
176	HEMCMV19	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	186	941	33	931	79	79	409	1	23	24	178
177	HEMDX17	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	187	654	1	654	137	137	410	1			13
178	HETAR54	209010 04/28/97 209085	Uni-ZAP XR	188	1848	454	1848	948	948	411	1	14	15	232

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
179	HETBX14	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	189	1146	157	1146		74	412	1	14	15	53
180	HFGAB48	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	190	906	156	906	245	245	413	1	30	31	32
181	HFKFI40	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	191	1941	120	1002	213	213	414	1	18	19	218
182	HFXHN68	209010 04/28/97 209085 05/29/97	Lambda ZAP II	192	2118	777	2118	966	966	415	1	23	24	50
183	HGBFO79	209011 04/28/97	Uni-ZAP XR	193	1538	259	1538	273	273	416	1	23	24	49
184	HGLAM56	209011 04/28/97	Uni-ZAP XR	194	1098	68	1098		185	417	1	28	29	69
185	HHLBA89	209011 04/28/97	pBluescript SK-	195	1001	1	1001	324	324	418	1	25	26	39
186	HHPDW05	209011 04/28/97	Uni-ZAP XR	196	1443	1	1443	246	246	419	1	21	22	21
187	HHPSD37	209011 04/28/97	pBluescript	197	1282	66	1282	171	171	420	1	19	20	37

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
188	HHP5F70	209011 04/28/97	pBluescript	198	951	26	951		162	421	1	16	17	34
189	HHSACK25	209011 04/28/97	Uni-ZAP XR	199	1740	1390	1740	1534	1534	422	1	19	20	31
190	HIASB53	209011 04/28/97	pBluescript	200	1707	401	1195	652	652	423	1	26	27	126
191	HJABZ65	209011 04/28/97	pBluescript SK-	201	779	1	779	23	23	424	1	26	27	68
192	HJPBB39	209011 04/28/97	Uni-ZAP XR	202	1617	188	1605	182	182	425	1	28	29	91
193	HLH5K94	209011 04/28/97	pBluescript	203	1974	1	1794	112	112	426	1	26	27	379
194	HLHTC70	209011 04/28/97	pBluescript	204	1057	229	1057	365	365	427	1	23	24	22
195	HLMIW92	209011 04/28/97	Lambda ZAP II	205	721	1	721	244	244	428	1	25	26	46
196	HLICY93	209011 04/28/97	Uni-ZAP XR	206	2465	988	2465	1225	1225	429	1			4
197	HLTDB65	209011 04/28/97	Uni-ZAP XR	207	1480	1	1480		371	430	1	15	16	143
198	HMSHM43	209011 04/28/97	Uni-ZAP XR	208	872	1	872	35	35	431	1	18	19	36
199	HMSHQ24	209011 04/28/97	Uni-ZAP XR	209	1779	16	1779	148	148	432	1	24	25	36
200	HNFAH08	209011 04/28/97	Uni-ZAP XR	210	2110	592	2110	611	611	433	1	18	19	191

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
201	HNGAO10	209011 04/28/97	Uni-ZAP XR	211	938	1	938	107	107	434	1	27	28	30
202	HNGBE45	209011 04/28/97	Uni-ZAP XR	212	1551	1	1551	114	114	435	1	21	22	100
203	HNHAZ16	209011 04/28/97	Uni-ZAP XR	213	997	1	997	202	202	436	1	24	25	36
204	HNHCM59	209011 04/28/97	Uni-ZAP XR	214	1496	1	1132		165	437	1	28	29	41
205	HOSFM22	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	215	1308	501	1308		809	438	1			1
206	HPHAC88	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	216	1705	384	1705	549	549	439	1	23	24	24
207	HCDEO95	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	217	999	608	999	273	273	440	1	22	23	54

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

5 The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, *Gene* 67:31-40 (1988).

10 Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

Signal Sequences

15 Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, *Virus Res.* 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, *Nucleic Acids Res.* 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1
20 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinje, *supra.*) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

25 In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., *Protein Engineering* 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results
30 shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., +
35 or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

10 Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

15 By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown in Table 1, the ORF (open reading frame), or any fragment specified as described herein.

25 As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are:
35 Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization

Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query

amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions,

5 interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be
10 determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and
15 subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window
20 Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity.
25 For subject sequences truncated at the N- and C-termini, relative to the the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of
30 the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are
35 considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level.

Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., *J. Biol. Chem.* 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after

deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon-gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

Polynucleotide and Polypeptide Fragments

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, or 2001 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity. More preferably, these polynucleotides can be used as probes or primers as discussed herein.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the

carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Particularly, N-terminal deletions of the polypeptide of the present invention can be described by the general formula m-p, where p is the total number of amino acids in the polypeptide and m is an integer from 2 to (p-1), and where both of these integers (m & p) correspond to the position of the amino acid residue identified in SEQ ID NO:Y.

Moreover, C-terminal deletions of the polypeptide of the present invention can also be described by the general formula 1-n, where n is an integer from 2 to (p-1), and again where these integers (n & p) correspond to the position of the amino acid residue identified in SEQ ID NO:Y.

The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini, which may be described generally as having residues m-n of SEQ ID NO:Y, where m and n are integers as described above.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

Epitopes & Antibodies

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an

epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')₂ fragments) which are capable of specifically binding to protein. Fab and F(ab')₂ fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the

polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

5 Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

10 Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final
15 preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

 Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins
20 facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG)
25 can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

 Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules
30 together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the
35 fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D.

Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol. Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein.

Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

15 **Vectors, Host Cells, and Protein Production**

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance

genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila* S2 and *Spodoptera Sf9* cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., *Basic Methods In Molecular Biology* (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein

after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

5

Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

10

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

15

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

20

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

25

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

30

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are

35

more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library) .) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model

systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erlich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of

unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell . Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (125I, 121I), carbon (14C), sulfur (35S), tritium (3H), indium (112In), and technetium (99mTc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, 131I, 112In, 99mTc), a radio-opaque substance, or a material detectable by nuclear magnetic

resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human
5 subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of ^{99m}Tc. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The
10 Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene
15 expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to
20 supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired
25 response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such
30 as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a
35 recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

Immune Activity

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can

decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

5 A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, 10 differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, 15 glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitis, and autoimmune 20 inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

25 A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The 30 administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may 35 inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic

shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1.)

Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstrom's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

Infectious Disease

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases

may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

5 Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, 10 Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., 15 Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiollitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, 20 Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that 25 can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Norcardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, 30 Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Meningococcal), Pasteurellaceae Infections (e.g., Actinobacillus, Heamophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, 35 and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS

related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria, Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas.

These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

Regeneration

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal

or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

Chemotaxis

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

5

Binding Activity

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit (antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

5 All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

10 Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with
15 a polypeptide of the invention, (b) assaying a biological activity, and (b) determining if a biological activity of the polypeptide has been altered.

Other Activities

A polypeptide or polynucleotide of the present invention may also increase or
20 decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic
25 surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, circadian
30 rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

A polypeptide or polynucleotide of the present invention may also be used as a
35 food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under
5 stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which
10 comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide
15 sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at
20 least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

25 A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide
30 sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer
35 as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method

comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

5 Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of
10 comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95%
15 identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

20 The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

25 Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous
30 nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

35 The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95%

identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide

comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

10 Examples

Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

	<u>Vector Used to Construct Library</u>	<u>Corresponding Deposited Plasmid</u>
	Lambda Zap	pBluescript (pBS)
	Uni-Zap XR	pBluescript (pBS)
	Zap Express	pBK
25	lafmid BA	plafmid BA
	pSport1	pSport1
	pCMVSPORT 2.0	pCMVSPORT 2.0
	pCMVSPORT 3.0	pCMVSPORT 3.0
	pCR [®] 2.1	pCR [®] 2.1

30 Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Altting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Altting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are
35 commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1

Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation of the fl origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the fl ori generates sense strand DNA and in the other, antisense.

Vectors pSport1, pCMVSPORT 2.0 and pCMVSPORT 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain

DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lacmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into E. coli strain XL-1 Blue. Vector pCR[®]2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into E. coli strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with ³²P-γ-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).)

The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 μ l of reaction mixture with 0.5 μ g of the above cDNA template. A convenient reaction mixture is 1.5-5 mM $MgCl_2$, 0.01% (w/v) gelatin, 20 μ M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., *Nucleic Acids Res.* 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then
5 be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA
10 synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

15
Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide

A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO:X.,
20 according to the method described in Example 1. (See also, Sambrook.)

Example 3: Tissue Distribution of Polypeptide

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by,
25 among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P³² using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is
30 then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are
35 mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

Example 4: Chromosomal Mapping of the Polynucleotides

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

Example 5: Bacterial Expression of a Polypeptide

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.⁶⁰⁰) of between 0.4 and 0.6. IPTG

(Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM imidazole. Imidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA

insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

- 5 The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

Example 6: Purification of a Polypeptide from an Inclusion Body

- 10 The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

- Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

- 20 The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

- 25 The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

- 30 Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

 To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area

(e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant A_{280} monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Commassie blue stained 16% SDS-PAGE gel when 5 μ g of purified protein is loaded.

The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

Example 7: Cloning and Expression of a Polypeptide in a Baculovirus

Expression System

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak *Drosophila* promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("Geneclean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five μ g of a plasmid containing the polynucleotide is co-transfected with 1.0 μ g of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., *Proc. Natl. Acad. Sci. USA* 84:7413-7417 (1987). One μ g of BaculoGold™ virus DNA and 5 μ g of the plasmid are mixed in a sterile well of a microtiter plate containing 50 μ l of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 μ l Lipofectin plus 90 μ l Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm

tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

- 5 After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.)
- 10 After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in
- 15 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

- To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection
- 20 ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of ³⁵S-methionine and 5 µCi ³⁵S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins
- 25 in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

30 **Example 8: Expression of a Polypeptide in Mammalian Cells**

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates

the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLV, HIV and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSPORT 2.0, and pCMVSPORT 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., *J. Biol. Chem.* 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., *Biochem. et Biophys. Acta*, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., *Biotechnology* 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., *Biochem J.* 227:277-279 (1991); Bebbington et al., *Bio/Technology* 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., *Molecular and Cellular Biology*, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., *Cell* 41:521-530 (1985).) Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the

polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five μ g of the expression plasmid pC6 is cotransfected with 0.5 μ g of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 μ M, 2 μ M, 5 μ M, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200 μ M. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

Example 9: Protein Fusions

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the half-life time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:

```

GGGATCCGGAGCCCAAATCTTCTGACAAAACTCACACATGCCCACCGTGCC
CAGCACCTGAATTCGAGGGTGCACCGTCAGTCTTCTCTTCCCCCAAAC
CAAGGACACCCCTCATGATCTCCCGGACTCCTGAGGTCACATGCGTGGTGGT
GGACGTAAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACG
GCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAAC

```

AGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTG
AATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAACCCCC
ATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGT
GTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCT
5 GACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATCGCCGTGGAGTGGGA
GAGCAATGGGCAGCCGGAGAACAACACTACAAGACCACGCCTCCCGTGCTGG
ACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCA
GGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGC
ACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGC
10 GACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

Example 10: Production of an Antibody from a Polypeptide

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

20 In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell
25 Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at
30 about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line
35 (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as

described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')₂ and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')₂ fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

Example 11: Production Of Secreted Protein For High-Throughput Screening Assays

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a

working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The
 5 PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2×10^5 cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x
 10 Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in
 15 Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of
 20 transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off
 25 PBS rinse, and person B, using a 12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of CaCl₂ (anhyd); 0.00130 mg/L
 30 CuSO₄·5H₂O; 0.050 mg/L of Fe(NO₃)₃·9H₂O; 0.417 mg/L of FeSO₄·7H₂O; 311.80 mg/L of KCl; 28.64 mg/L of MgCl₂; 48.84 mg/L of MgSO₄; 6995.50 mg/L of NaCl; 2400.0 mg/L of NaHCO₃; 62.50 mg/L of NaH₂PO₄·H₂O; 71.02 mg/L of Na₂HPO₄; .4320 mg/L of ZnSO₄·7H₂O; .002 mg/L of Arachidonic Acid ; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic
 35 Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of Oleic Acid; 0.010 mg/L of Palmitic Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of

Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L- Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine-H₂O; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCL-H₂O; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0
 5 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL-H₂O; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-Lysine HCL; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalanine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tyrosine-2Na-2H₂O; 99.65 mg/ml of L-
 10 Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319 mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B₁₂; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine;
 15 0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mm glutamine and 1x
 20 penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B
 25 adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

30 It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an
 35 activity in a particular assay.

Example 12: Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

	<u>Ligand</u>	<u>tyk2</u>	<u>JAKs</u> <u>Jak1</u>	<u>Jak2</u>	<u>Jak3</u>	<u>STATS</u>	<u>GAS(elements) or ISRE</u>
	<u>IFN family</u>						
5	IFN- α /B	+	+	-	-	1,2,3	ISRE
	IFN-g		+	+	-	1	GAS (IRF1>Lys6>IFP)
	IL-10	+	?	?	-	1,3	
	<u>gp130 family</u>						
10	IL-6 (Pleiotrohic)	+	+	+	?	1,3	GAS (IRF1>Lys6>IFP)
	IL-11(Pleiotrohic)	?	+	?	?	1,3	
	OnM(Pleiotrohic)	?	+	+	?	1,3	
	LIF(Pleiotrohic)	?	+	+	?	1,3	
	CNTF(Pleiotrohic)	-/+	+	+	?	1,3	
15	G-CSF(Pleiotrohic)	?	+	?	?	1,3	
	IL-12(Pleiotrohic)	+	-	+	+	1,3	
	<u>g-C family</u>						
20	IL-2 (lymphocytes)	-	+	-	+	1,3,5	GAS
	IL-4 (lymph/myeloid)	-	+	-	+	6	GAS (IRF1 = IFP >>Ly6)(IgH)
	IL-7 (lymphocytes)	-	+	-	+	5	GAS
	IL-9 (lymphocytes)	-	+	-	+	5	GAS
	IL-13 (lymphocyte)	-	+	?	?	6	GAS
25	IL-15	?	+	?	+	5	GAS
	<u>gp140 family</u>						
	IL-3 (myeloid)	-	-	+	-	5	GAS (IRF1>IFP>>Ly6)
	IL-5 (myeloid)	-	-	+	-	5	GAS
30	GM-CSF (myeloid)	-	-	+	-	5	GAS
	<u>Growth hormone family</u>						
	GH	?	-	+	-	5	
	PRL	?	+/-	+	-	1,3,5	
35	EPO	?	-	+	-	5	GAS(B-CAS>IRF1=IFP>>Ly6)
	<u>Receptor Tyrosine Kinases</u>						
	EGF	?	+	+	-	1,3	GAS (IRF1)
	PDGF	?	+	+	-	1,3	
40	CSF-1	?	+	+	-	1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:

5':GCGCCTCGAGATTTCCTCGAAATCTAGATTTCCTCGAAATGATTTCCTCG
AAATGATTTCCTCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTTGTCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

5':CTCGAGATTTCCTCGAAATCTAGATTTCCTCGAAATGATTTCCTCGAAATG
ATTTCCTCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCC
CTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGC
CCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGC
CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTT
TGCAAAAAGCTT:3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, IL-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

Example 13: High-Throughput Screening Assay for T-cell Activity.

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1%Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies)

with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells (10^7 per transfection), and resuspend in OPTI-MEM to a final concentration of 10^7 cells/ml. Then add 1ml of 1×10^7 cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Gentamicin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100,000 cells per well).

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophane covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

Example 14: High-Throughput Screening Assay Identifying Myeloid Activity

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells.

- 5 Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

- 10 To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest 2×10^7 U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

- 15 Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, 1 mM MgCl_2 , and 675 uM CaCl_2 . Incubate at 37°C for 45 min.

- 20 Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

- 25 These cells are tested by harvesting 1×10^8 cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of 5×10^5 cells/ml. Plate 200 ul cells per well in the 96-well plate (or 1×10^5 cells/well).

- 30 Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat phenochromocytoma cells) are known to proliferate and/or differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)

5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine

growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as 5×10^5 cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to 1×10^5 cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

Example 16: High-Throughput Screening Assay for T-cell Activity

NF- κ B (Nuclear Factor κ B) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- κ B regulates the expression of genes involved in immune cell activation, control of apoptosis (NF- κ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- κ B is retained in the cytoplasm with I- κ B (Inhibitor κ B). However, upon stimulation, I- κ B is phosphorylated and degraded, causing NF- κ B to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- κ B include IL-2, IL-6, GM-CSF, ICAM-1 and class 1 MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- κ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- κ B would be useful in treating

diseases. For example, inhibitors of NF- κ B could be used to treat those diseases related to the acute or chronic activation of NF- κ B, such as rheumatoid arthritis.

To construct a vector containing the NF- κ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- κ B binding site (GGGGACTTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:
 5':GCGGCCTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGAC
 TTTCCATCCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:

5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene)

Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5':CTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGACTTTCC
 ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCA
 TCCCGCCCCTAACTCCGCCCGAGTTCGCCCATTTCTCCGCCCCATGGCTGACT
 AATTTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTC
 CAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTT:
 3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2- promoter plasmid (Clontech) with this NF- κ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- κ B/SV40/SEAP cassette is removed from the above NF- κ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the NF- κ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF- κ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

Example 17: Assay for SEAP Activity

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15 μ l of 2.5x dilution buffer into Optiplates containing 35 μ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50 μ l Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50 μ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4
15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6

23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO₂ incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 μ l of 12 μ g/ml fluo-3 is added to each well. The plate is incubated at 37°C in a CO₂ incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 μ l of buffer.

5 For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2.5×10^6 cells/ml with HBSS in a 50-ml conical tube. 4 μ l of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1×10^6 cells/ml, and dispensed into a microplate, 100
10 μ l/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 μ l, followed by an aspiration step to 100 μ l final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the
15 following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 μ l. Increased emission at 530-nm indicates an extracellular signaling event which has resulted in an increase in the intracellular Ca⁺⁺ concentration.

20

Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase
25 RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

30 Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members
35 of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford, MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford, MA) are used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na₃VO₄, 2 mM Na₄P₂O₇ and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a

biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

- 5 The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg₂₊ (5mM ATP/50mM MgCl₂), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl₂, 5 mM MnCl₂, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the
- 10 components gently and preincubate the reaction mix at 30°C for 2 min. Initial the reaction by adding 10ul of the control enzyme or the filtered supernatant.

 The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mM EDTA and place the reactions on ice.

- Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction
- 15 mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as
- 20 above.

- Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of
- 25 tyrosine kinase activity.

Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity

- As a potential alternative and/or compliment to the assay of protein tyrosine
- 30 kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase,
- 35 Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other

phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies).

The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

5 PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

10 Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenindeoxy-uridine 5'-triphosphate (Boehringer Mannheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

15 Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated
25 disease.

Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

30 A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

35 For example, antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10.

The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

Next, 50 μ l of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

Add 75 μ l of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

Example 23: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 μ g/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μ g/kg/hour to about 50 μ g/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally,

intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules.

Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., *Biopolymers* 22:547-556 (1983)), poly (2- hydroxyethyl methacrylate) (R. Langer et al., *J. Biomed. Mater. Res.* 15:167-277 (1981), and R. Langer, *Chem. Tech.* 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., *Proc. Natl. Acad. Sci. USA* 82:3688-3692 (1985); Hwang et al., *Proc. Natl. Acad. Sci. USA* 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 24: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form.

- 5 Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

- 10 For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

Example 25: Method of Treating Increased Levels of the Polypeptide

- 15 Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

- For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 20 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

Example 26: Method of Treatment Using Gene Therapy

- 25 One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is 30 turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

5 pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

10 The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to
15 transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is
20 then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media,
25 containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is
30 required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 27: Method of Treatment Using Gene Therapy - In Vivo

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense
5 DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide of the present invention. A polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the encoded polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art,
10 see, for example, WO90/11092, WO98/11779; U.S. Patent NO. 5693622, 5705151, 5580859; Tabata H. et al. (1997) Cardiovasc. Res. 35(3):470-479, Chao J et al. (1997) Pharmacol. Res. 35(6):517-522, Wolff J.A. (1997) Neuromuscul. Disord. 7(5):314-318, Schwartz B. et al. (1996) Gene Ther. 3(5):405-411, Tsurumi Y. et al. (1996) Circulation 94(12):3281-3290
15 (incorporated herein by reference).

The polynucleotide constructs of the present invention may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). These polynucleotide constructs can be delivered in a
20 pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the
25 polynucleotides may also be delivered in liposome formulations (such as those taught in Felgner P.L. et al. (1995) Ann. NY Acad. Sci. 772:126-139 and Abdallah B. et al. (1995) Biol. Cell 85(1):1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs of the present invention used in
30 the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the
35 transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct of the present invention can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 g/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for the polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with

liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e.g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice. The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA of the present invention.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

Sequence Listing

(1) GENERAL INFORMATION:

- 5 (i) APPLICANT: Human Genome Sciences, Inc., et al.
- (ii) TITLE OF INVENTION: 207 Human Secreted Proteins
- 10 (iii) NUMBER OF SEQUENCES: 800
- (iv) CORRESPONDENCE ADDRESS:
- 15 (A) ADDRESSEE: Human Genome Sciences, Inc.
- (B) STREET: 9410 Key West Avenue
- 20 (C) CITY: Rockville
- (D) STATE: Maryland
- (E) COUNTRY: USA
- 25 (F) ZIP: 20850
- (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
- (B) COMPUTER: HP Vectra 486/33
- 35 (C) OPERATING SYSTEM: MSDOS version 6.2
- (D) SOFTWARE: ASCII Text
- 40 (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
- 45 (B) FILING DATE:
- (C) CLASSIFICATION:
- 50 (vii) PRIOR APPLICATION DATA:
- (A) APPLICATION NUMBER:
- 55 (B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Kenley K. Hoover

(B) REGISTRATION NUMBER: 40,302

(C) REFERENCE/DOCKET NUMBER: PZ007PCT

(vi) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (301) 309-8504

(B) TELEFAX: (301) 309-8439

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 733 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGGATCOGGA	GGCCAAATCT	TCTGACAAAA	CTCACACATG	CCCACCGTGC	CCAGCACCTG	60
AATTCGAGGG	TGCACCGTCA	GTCTTCCTCT	TCCCCCAAA	ACCCAAGGAC	ACCTCATGA	120
TCTCCCGGAC	TCCTGAGGTC	ACATGCGTGG	TGGTGGACGT	AAGCCAAGAA	GACCCTGAGG	180
TCAAGTTCAA	CTGGTACGTG	GACGGCGTGG	AGGTGCATAA	TGCCAAGACA	AAGCCGCGGG	240
AGGAGCAGTA	CAACAGCACG	TACCGTGTGG	TCAGCGTCCT	CACCGTCCTG	CACCAGGACT	300
GGCTGAATGG	CAAGGAGTAC	AAGTGCAAGG	TCTCCAACAA	AGCCCTCCCA	ACCCCCATCG	360
AGAAAACCAT	CTCCAAGCC	AAAGGGCAGC	CCCGAGAACC	ACAGGTGTAC	ACCCTGCCCC	420
CATCCCGGGA	TGAGCTGACC	AAGAACCAGG	TCAGCCTGAC	CTGCTGGTTC	AAAGGCTTCT	480
ATCCAAGCGA	CATCGCCGTG	GAGTGGGAGA	GCAATGGGCA	GCCGGAGAAC	AACTACAAGA	540
CCACGCCTCC	CGTGCTGGAC	TCCGACGGCT	CCTTCTTCCT	CTACAGCAAG	CTCACCCTGG	600
ACAAGAGCAG	GTGGCAGCAG	GGGAACGTCT	TCTCATGCTC	CGTGATGCAT	GAGGCTCTGC	660
ACAACCACTA	CACGCAGAAG	AGCCTCTCCC	TGTCTCOGGG	TAAATGAGTG	CGACGGCCGC	720
GACTCTAGAG	GAT					733

(2) INFORMATION FOR SEQ ID NO: 2:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

- 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
Trp Ser Xaa Trp Ser
1 5

- 15 (2) INFORMATION FOR SEQ ID NO: 3:

- 20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 86 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
25 GCGCCTCGAG ATTTCCCGA AATCTAGATT TCCCCGAAAT GATTTCCTCG AAATGATTTC 60
CCCGAAATAT CTGCCATCTC AATTAG 86

30

- (2) INFORMATION FOR SEQ ID NO: 4:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

- 40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
GCGGCAAGCT TTTTGCAAAG CCTAGGC 27

45

- (2) INFORMATION FOR SEQ ID NO: 5:

- 50 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

- 55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
CTCGAGATT CCGGAAATC TAGATTTCCT CGAAATGATT TCCCCGAAAT GATTTCCTCG 60
AAATATCTGC CATCTCAATT AGTCAGCAAC CATAGTCCCG CCCCTAACTC CGCCCATCCC 120
- 60

180
GCCCCTA ACT CCGCCCA GTT CCGCCCA TTC TCGCCCA CAT GGCTGACTAA TTTTTTTTAT
240
TTATGCAGAG GCCGAGGCCG CCTCGGCCTC TGAGCTATTC CAGAAGTAGT GAGGAGGCTT
5 TTTTGGAGGC CTAGGCTTTT GCAAAAAGCT T 271

10 (2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
15 (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

20 GCGCTCGAGG GATGACAGCG ATAGAACCCC GG 32

25 (2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
30 (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

35 GCGAAGCTTC GCGACTCCCC GGATCGGCCT C 31

40 (2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
45 (B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

50 GGGGACTTTC CC 12

55 (2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 73 base pairs
60 (B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

5 GCGGCCTCGA GGGGACTTTC CCGGGGACTT TCCGGGGACT TTCCGGGACT TTCCATCCTG 60
CCATCTCAAT TAG 73

10

(2) INFORMATION FOR SEQ ID NO: 10:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 256 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CTCGAGGGGA CTTTCCCGGG GACTTTCCGG GGACTTTCCG GGACTTTCCA TCTGCCATCT 60
25 CAATTAGTCA GCAACCATAG TCCGCCCCCT AACTCCGCCC ATCCCGCCCC TAACTCCGCC 120
CAGTTCCGCC CATCTCCGC CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA 180
GGCCGCCTCG GCCTCTGAGC TATTCCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCCTAGG 240
30 CTTTTGCAAA AAGCTT 256

35

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2526 base pairs
40 (B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

45 GACAGGCTAT CCGAGAATCT GAGAGCTGGG CCGGCAATT CCTCCAGYTA CCTTGTGAC 60
CTAAGTCCAG TCACACATTT CCCAAAGTTT CTCTTTGTCA TAACCCCTGGT CTGGCTGGTT 120
50 TTGRGGRCTT GAGAATGGGT CAGGGACTCC AGGCCAAGTC CAACAGAGAC CCCAAACCCA 180
CCACACACCA GCAGCCACAA CCTCACCACC AACAAAGAGG ACTTTTGTGG GGCCACAAGT 240
AAGAGGTCAT TTCTGGAATG GACTCAGACC TTAAACAGG AGAGTTGAGC ACTTCCAGKS 300
55 AGTTTTTAAG CAAGGCATGG GGAACAGGGA ATAGAACCTT TCAAAGAGGT TGCCAGAGA 360
AAAGCTGGGC CTCTTGCAAT CGGCTTCCTT GGAGCAGCCT CTCTTGCCAG AAAGCCATCA 420
60 GGTGCTCAAT CATCTTCTCC TGGCCAAGGC TCTGACCATG CTTAGTACTG GAATAGAGGT 480

	GCCCAGGCC CCAGCGACTC TTCTTGGCCT GATGTTTGTG CTCACAGGCA TGCCACGTGG	540
	CCTGAGATGA TTCAGAACAA ATCATGCTAA CTTTGAATCC ATCCAGCCAC TTGCAAATGA	600
5	TAATCAGAAG TCAGCTTGTT CACTGTTAGA AAGAACTAA CAAAAGAGAA CCCAGAGCAA	660
	TCTAGAATCT TTGAGTGCCT GGCTTTCCAA GGATACTGCG GAGACTCTGG CCAAGCTGAT	720
10	GAMCTTCTGA ARTGTCACTG GCACCATATG CAACAAGAAC CACCATTAC TGAGTAGCTA	780
	ATGGGTTTGG GGCCTGGGAC ATTCCATCTG AGGTCCTTCC TGAACATGTC ACTCCACAGC	840
	AGAGGACCGG TTGCAGCTTA CCCAGAACCA CTCCTCCAGG AGAGCTGGAT GTTTTGCGTG	900
15	CAACACCTTG AGCACTGACT GCTATTGTTC AAAAAAGCC TTGCTGCAT TCGGAGGACT	960
	GCCCCGTGCC CTGAGGTGAC TTCCTAACTA TGTGGTTTCA TTAGCGAATT TATTTTTTGT	1020
20	GCTGGGTGGA CATTTGTATT TTGTTAGGTT GCTGTTTAAG CTCAAGTTTG CTGTGCTCTC	1080
	TGCAGCTACA AACATCTTG GCATATTTAA GAKTGGCTTT TATAAATAGC TTTATTCTGA	1140
	TATTAATCAG ATTCCCAACT TTACTGAGAA TTAAGGACTG GGTACTTTA AAGAAATGCA	1200
25	AATAGCAATT GAAGAACCAC TGCTGCAGGT GGTAGCCCTG GCTAGACTGA ATTACACTAG	1260
	AAATCAGCCA GAAGGAAGCG TCCTTGGGAT CCCAGATCAC TCTTTTTTTT TTTTTTTTTA	1320
30	AAAGGGGAG CCCCTTGATG GCTCATCTCT CTGAATAACA GTTACGTCCT CATATCGATA	1380
	CCAGATGCCT TCTTCATCAT GCCACTGAAG CCACTCACCA CCTTCAAGAA CATGCCAACC	1440
	TCTGTCAGAT TCACTTACCC ACAACAAGG AGGCACGTTT GGCACAAAGT GTTGTCTCTC	1500
35	AGGTCCAAGT GGAATCTACA GAGTGTCTGA CCTCAACACA CTGGATTCCA GGTGGACTGG	1560
	ACCAAGAGCA GGCAAGACA CGGGAAGTGA AAAACTCCAC AGGGTTTGA GAATAGAAAT	1620
40	GAAAAGCCAC GTCATATAAC TCAAGAATAA ATGGTGTCTT GGAAATTTTA AAATTATCAT	1680
	CGAAGGTGGT GAAACTATTT CAGGCCCAAA TGAAAGGAAA TCGCCAGTTG GGGATGAAAT	1740
	CACAGAGCCT GTGTTTATG ATATGGTTGG ATGTCCACTG ATGAAATTTT AAAGGAGTTT	1800
45	CATTTTTTAA AGTGGCATG ATTCTACATA TGAGAATTCT TTAGGCCAAG AAATGTCTCT	1860
	TGGCTCAGAG GTGTTGGGAA TTAAAGCAGA GAGAAGCCAT TCGTGATGCT TAGAACCAAG	1920
50	GATGGTCATG TACACAAAGA CCATCGAGAC GCCATTCTT GTTTACAAAA CACTTACCAA	1980
	GAAAGCACTT TGTAGGGGAA CTTTAGTAAG TTCTTCTCAT TTCATTATGT TTCTTCCAAG	2040
	GAAACAGGAG AGACTGAATT AATAATTCTC TCTTCTCTCT TAAGCACTTT TAAAATAATA	2100
55	AAGTACATCT TGAAATTTGG GGGGCATCT CTGATTTAAA AAAAGAAAAA GGCTGCTTGA	2160
	TGTATGTTAT GCAGAGACAC TCTGCCTCTG GTGGCTGCAG AGCAATACCC AAGCCTCAT	2220
60	TGGAAGGCTC AACATTTGGA ATTGCATTTT AATTGATTAA TCCTCAATTC ATGTGGCCTT	2280

ACGGGATGGT GGGTCTGGGA CCCCAATTCA TTCTTATCTG CCAAAGAATT ATCTAGAAGC 2340
 ACATCAAATA CCAGCACCCC ACCTGCACAA TGGGGGTGGA AAACITTTTGT ATCCCTAAGC 2400
 5 ATATTATTTT ATAGTGTCTG CCATGCCATG TGGAAATACT TTATTTTAA CCTCAGGATT 2460
 TAAATAAAGT AAACACTATG ACATTTAAAA AAAAAAAAAA AAAACTCGAG GGGGGCCCGG 2520
 10 TACCCA 2526

15 (2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1131 base pairs
 (B) TYPE: nucleic acid
 20 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

25 CACTGCACCA GCTTTGTTAT CTGTAAAATG ATGATAATAC CAACACCTTC TTCTTGGGGT 60
 ACTGAAGATG AGAGAACATG ATATGTGTAA AGTGCCCTCC ACAATACCCA GAACATAGCA 120
 AACATGTAAT GAATGTAGTA ATAGTAATTA TTTATTTTC TTTTGATTCA GTTGGGACTA 180
 30 TGTTCACTG TAACAGAATA CCCAAAATAA CTGTTTTTAA CAAATTAAG TTTWGTGTG 240
 AAGTTTTGTT ACGAATTCAG ACAATCCAGG GCTTTTATAG ATGCACCAGG ATCAGCAGGT 300
 35 ACAAAGGCAT CTTCCTGAT TTCTGCCAGT CTCAATGCAT GGGTTGCAAT CCAGATCCA 360
 RGATGGCAGT TCCAGCCCTG GTTAGGCCCA TATTAGCACA CAGAAAGAAA GAGAAAGGGA 420
 TGTGCTCTTT CACTTTAATC ATAGCTCCCA CTAGATGCAC CCACTACTTC TGCTGATACT 480
 40 CCATTAGCTA ATGCTTGCTT ACATGGTCAC ACTTAGTTTC CAGAGAGACA TGTCTGGACA 540
 GTCATGTGCT CAATTAATAT CCAAGTGTCC AATTACTGAG AAAAAAGAA ACTAGCACCT 600
 45 TTGCTTGCTT GCATTCCTCT TAGCATAAGC CACATCTCTT TTATGAAGTT GTCCTCAGTT 660
 ACTTGGATGC CTCAGTTGTC CTTTCAWTTA GAAAWGCYCC TKGGACAYCC TGAAWCTGAC 720
 TTCITTTGTC ATCAGCACCA TCACTACCAC TGCCYTCTTC AAAGCCACCA CGTTCTGTCC 780
 50 CCAGGATGGT TGCAACAACC ACCATAGGGA CTTTTTGCCT TCTACTTCCA CACAATAGNC 840
 CAGAGTAAGC TTTTGAAAT GTAGGTCAGA TCATGTCTCT CTCTCTCTCT TCAAAACCCT 900
 55 CCCGATGGCT TTTCATATTA CTCAAAAGAA AACCTAAAAC TTTGCTGTGA GATCTATGTG 960
 ACCCGGCTTA TTCTTCTCT TACTTTATCT CTGTATGCT CTCTCTCACT CTACTCCAGC 1020
 CATCCACCT CTTGCTGCT TGTCTATAC TCCTAAAAGA AGTTCAGTCT TCCCTTATGA 1080
 60

TATTTGCACT TAAATAGAA AAAAAAAAAA AAAAAAACT CGAGGGGGGC C

1131

5

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 941 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

15

GGCACAGTA GCATTTCACT TAATCTGCAG GTATATTCTC CCAACAGTTT ATTGTCATGT 60

GATGTCCTCA GCCAAGATTG TRAGGCAGAG AGGAGCTGTC CCAACCTACT ATACCACCGA 120

20

GGCTGGAGAG ATCATATTTT TGGTATTAAA CTGGAGTCTC TCCATCCTTC ACATTGTTGA 180

TGTCCTCTGT AGCAAACCGG AAAAGTCAGT GACAGAAGAT GCCGCTAGCG GTTTGAGCCA 240

GAGAAATGACA GCTCTGGTTT GGAGAAAAGG GCCGGATGGT GGCTCTAGAA AGCCCATCCT 300

25

TCTGCTCTTC TTTTCTCTCC CCTTATATT GTGCTTTCAT TCATTCATTC ATTATCAAA 360

CATTTGTTGA GCACCTATTA TGTGTCAAGC TCTGTGCTAG CCTCTGGAAA ACCTGCCCTC 420

30

ATGTAGCTCA CTGTGGAGTA GGAGAAACAA TGACTACACT ATGATAAGCA CGGGTTGTCA 480

GGGTCTCACA GAGCAGTGGC CCTCATCCA GACCGATGAG GTCAAAGAAG GCATCCAGGC 540

GAGGATGGTG TCAGAGCTAA CTGAAGAATG AGAGGGAGCT GCACCASCAG GGGTTGGAAC 600

35

TGAAGGTGGC AGTGCCTGGA GTCTTGATTC CAGCAGAGGG AGAGCAGTCT GTGAAAAGGC 660

ACCAAGGGTG GGAGAGGGCA GAGCACATGG AGGAACTTCA GGTAGTTCTG GATGGCCTG 720

40

GGGCAAAGCT AGAGAGGTAA GAAGAATCTA CAAATGTTC TCGAGTTACA TGAACCTCCA 780

TCCAATAAAA CCATTTGGAA ACGAAAAATT TAAGTCAGAA GTGCATTAA GGCTGGTCCG 840

AGTAGAATGA TTTTACAAC GAATTGATCA CAACCAGTTA CAGATGTCTT TGTTCCTTCT 900

45

CCACTCCAC TGCTTCACCT GACTAGCCTT TAAAAAAA A 941

50

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

55

- (A) LENGTH: 843 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

60

CNAGGGATAA CCCCAAAGNT GGGAAATAAA CCCTCAATTA AAGGGGAAC CAAAAAGCTG 60
 GGAAGTTCCC CCCCGCGGTG GCGGCCNGNT CTAGGAATA GTGGAATCCC CCGGGGCTGC 120
 5 AGGGAATTCG GCACGGAGTG GGAATGTTGT TTGTATGATA CTATTTCCAC AAWATGCATT 180
 GAGACTTGGT KTGTGGCCTA GGACATGGTC AATTCTTTT AAATATTCCG TGAATTTCTT 240
 10 TAGTGCATAT TCTCCGATGG GGGCTGTGGG GACAGAGTTC TAAATATGCC CATTAGATTA 300
 AATCTCTTCA TTCTGTTGCT CACATCTTCT ATATCCTTAT TAATCTGTCA ATCTCTTCAA 360
 GAGAGGTGTT ATTAAAATCT CTCACTGTAT GTGTCACTTT GCCCTTAAAA TTCTGATGAT 420
 15 TTGCTTTATA AATGGTTATA ACCATTTTCC AGGAAGAACA TTAAAGAACT TTCCATTGGC 480
 ATTATCCAGT TTCCCTCAAA ATACTGGTTT TTTTATTTT GGCTNCTAAG CAGCTATGAA 540
 TCCAGTTTCT CAGAAGCCCT TGTCTCAAGG CATTTGTTTC CAGATTACCT TGTTAGCATC 600
 20 CACACTATGG GCTATTTTAG AAAACAAAA AAAGTATCAA AATCATATAG CTATGATTTT 660
 CCTGTGCTTG AAGGAGCCTT AAAGCTCATC TAGTCCAGCC AGTATTTGTT CATCCAAATT 720
 25 CTGCCAAGAA ATCTCTATTG TCAAGATATT CTTTACCATC TTGGGACAT TCTCATTATT 780
 AGAAACAAAT CCTAAGAAGA AATTCTGCCA TAKACAACCC ATCCGTCTTT TAAAAAATAA 840
 AAA 343
 30

(2) INFORMATION FOR SEQ ID NO: 15:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1018 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CTGTAATTTT TAATTTTCAT ATACCGTGCT TTGATTCTAA TTTTATTTT TGAGTTCTCT 60
 45 GAAGGTACATA TATACAGAGT GCTTCAGGAA TGATCATTTT GTTATTATTC ATGCTTCTTA 120
 ACAATGTTGT TTTAGTCCAA GAAGATAATT GCCAGAGAAA GAATACAGTG CAGGAAAGAA 180
 50 GARGCTGGAG CCACTGGTGA AGARGGATTG AGARGACAGA CATTTGTGGGA ATGAAATCAT 240
 GAATAATCGT GTTTTGAAT TGTCCAAAAA CTCTACAAA CCATGAAATG TTGGAGTTTA 300
 AATCTAATTG TTGAAAAATT CCCACATTC CTGTATCCC TTAGGTTGAG CATAATTCCA 360
 55 CATCCGTGGA CTGATGCACT TCCCAAGAGG GGGCTCATT AACTCTTCCG AGGCAGCAGC 420
 AGCAAGGGCA CCCCTCCTT TCCCCCACA CCCAYTTCT CATGGCTCTT CTCTCTCTCA 480
 60 TCTCATGCTT AGGTTAGAAA AGGGCACAAG GTAAGGAAGC CCTTGGGAAT AGGCTGAATC 540

5 TGGCTATCTA ATTTGGTGCC AAATACTTAA TGTGCTTGAA TTTAAAAACA GCAAACATGT 600
AGAAAGGTAA TTATAATTAT GAGGCCAGTT CTTTAAGCTA GCTTTTTTTC CCTCTCAAA 660
CAGCATATTG GCTTGGATGT CAGCAGGAGA AAGTGTMTT TGCAATACAC ATAATGCATA 720
TATGGTCCTG TTAGCAATCT ATAGAAAATA GATATTGCTC ATTAAGGTAA ATATTTTGT 780
10 TGATGAATGA TCTGGAATGG TCTGGACTTG TTGTGTGAAC AGGAAATTGC TCTGTAGGCT 840
TTGACTTGTG AGGTAAAGAG TGAGGCTGGT AAGATTAATT AAAGTAAATA CTGTGACAAT 900
AGGATGTCAA AACCAAAAAC GTGTTTCTGA AACTCAAGGA ATTAATGACA CATAGGGAAG 960
15 TTTTGGCCAT ATTAAGCATA GAGTAGGAGA GGCAAGTCAA GAATAAAAAA AAAAAAAA 1018

20

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
25 (A) LENGTH: 661 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TTTAAGAAAT TAGTGAATCC CCGGNTGCAG GGAATTGGC ACGAGGAGGA GGCCGTCAGC 60
TGCCAGGAGC GCAGGATGCC AGCTGYTCCC CCGGGTGTGA CCCCCCAGY TCTGCTGGAC 120
35 ATAAGYTGGT TAACAGAGAG CCTGGGAGCT GGGCAGCCTG TACCTGTGGA GTGCCGGCAC 180
CGCCTGGAGG TGGCTGGGCC AAGGAAGGG CCTCTGAGCC CAGCATGGAT GCCTGCCTAT 240
GCCTGCCAGC GCCCTACGCC CCTCACACAC CACAACACTG GCCTMTCCGA GCTGCTGGAG 300
40 CATGGAGTGT GTGAGGAGGT GGAGAGAGTT CGGCGCTCAG AGAGGTACCA GACCATGAAG 360
GTGCGCAGGG CAGGGCTCGG ACCTACCCCA GGAATGTCTT GCCCTGGGAA TGACAACACA 420
45 GTCCACACCA TGCACGGGGA GGCAAACAGG GGCAGCTGAC CCAGCCCAGG GGTGAGANGA 480
GGTCTTGCCG AGGAAGTGGC AGCTAAGCTG ATACCTGATA TGCACWAGKC AGCCARGYGG 540
AGACAGGCAA GGAAGAAGCT TGTTTGTAGG ACAGAATTTT CTAGATCACT CAGCACCATC 600
50 TGGCTTTTGG GGCTTTTGT TTTATTTTGT TTTTGAGACG GGGTCTGCT CTGTGCCCCA 660
N 661

55

(2) INFORMATION FOR SEQ ID NO: 17:

60 (i) SEQUENCE CHARACTERISTICS:

275

- (A) LENGTH: 553 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

	GGCACAGGGC TATTTGCCCC TCTCTCCACA TGACAGAACT GCTCTAAGTT TCTTTGCTGC	60
10	TCTTCTCAGC TGTCAGACGG CTGCTGCTT GTTTTCCACA CCACCATGTC TATCTTTGC	120
	TGTCCTWAC TCTGCCTGTT TTTTCCCTT TGTATTTCTT CTGGCTCTTG TCCCTTTTCC	180
	CACGTGTCWC AGCTTTCTCT TATTGCCACT TTCAGTCAGA GCAGTCTGT GCTTCTGGTG	240
15	CCGGCATACA ATACTTACTT GAGTTTCTTG GCTTTTCTTG ACTGTGCATC TCTTACTTCA	300
	ACATAGGAAT AGCCTGTCAT AGAATTTCTC CAGTTCAGG GCTCAAGAGG GAGAGTGCCA	360
20	GAAAATTGAG ACTGTTTTCC CTGTCTTGA TTGAATTCAT AAAGCAAAC CAGTGTGTGT	420
	GTGAGGGTTT GCTGTGTCAT GCCTATAGGT TGTTTGGGTG CAAACCTATA GAATCCAGCC	480
	TGCGAAAAGA AAGRAACCAG AGAATANCAG CATCAGAACA ATGCTTGACA TCATTTCTCA	540
25	ATCAAGCAGT CCA	553

30

(2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 869 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

40

	GGCAGAGCT GCCAACACTG AGGTCTTCGT GGCTTCTCAC ATCTAGATGT ATCCCTCTCA	60
	AATCTATCCT CTATCCAGGC ACCAGATTGA GGTATCTAAA ATGTCAACTT TCCAGTTACT	120
45	CCTTCTTATA CTAGCCCAAT CAACTTACAA GATAAAGTCC AAGCCCTTC ATATGACAAA	180
	CCACACCCTG CTTAACCTCT CAGGTTTGAA TCCTTCATCT CTTACTTTAA ACTTTAAAC	240
	CCAGCAGCAC GAAAGTGTCT CCTATGCATG TTGCCATATG CGTTCTCTCC ATCATGCATT	300
50	TGCCTGAGCA AGATGTCTTG AGTTAACATC TTATTCITTA AGACTCATTG TGGTGGTAGA	360
	CAGCCTTTAA TAACGGATCC TTGGCCAGGC ACAGTGACTC ACACCTGTAA TCCCAGAACT	420
55	TTGAAAGGCC AAAGAAGGAA GAAAGCTTGA GGCCAGTAGT TTGAGACCAG CCTGGGAAAC	480
	AGAGAGATAT CCCATCTGTA CCAAAAATTT AAAAAATAT TAGCAGGGAG TAGTGGCATG	540
	CACAAGTGGT CCCAGCTCCA TGGGAGASTG AGGTAGGAAC ATCACTTGAG CCCAGGAAGT	600

60

276

CAAGGCTGCA GTGAACCATG ATCAGAACAT TGCANTCCAG CTGGGTAAC AGAGTGAGAC 660
CTTAGGTCAG AAAAAATGAAT AAATAAGCAT AAAATTTTAA AACTTAGCC AGGCATGGTG 720
5 GCACACATCT GTGGTCCCTG CTACTTAGGA GGCTGAGGTG AGAGGATCCT TGAGCCCAGG 780
AGGTCAACAC TACAGTGAGC TATGATTGTG CCACTAAACT CCAACCTGGG TGAAAAAGCA 840
AAACCCCTGCC AAAAAAAAAA AAAAAAACT 869

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 959 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

GGCGAGCCGA GATCGTGCCA TTGCACTCCA GCTGGGCAA CAAGAGTGAA ACTCTGTCTC 60
25 AAAAAAAAAA AATTATAATA CTATATGCCA TAAATGACA TTTCATATT AAAGAGTTT 120
TTAAACTCT TGTATTACA TGCCATAAT TGAAACCTA TTTCAGTAA TGAGAAATGGT 180
30 ATCTGTGTG CTCAITTTTT CATTTTATC CTTAACAATT TCCACCACAG CCAGTGCATA 240
TAATGGCAAT GACACCCAGG GATGGAATGA TAAGTTCCAT CRCMGCTCAG TCAAGACGCA 300
GACTTGATGT GGCCCAACA ACAGTCAATA ATGGAGTCTC CAAATAAAG CTCTATAGGA 360
35 AAGGTAAATA CCCGCTGCAC AAGAAACCAC AGCATCTAGG TTCTAACCCC ATCTCTATGA 420
AGAGCTTGCT GGGAGAGTTT TGACATTWAA CAATCTGTCT GATKGCCAAT TTTTTCCTC 480
40 TATAAAATGA TAATGTTKGA YTCAAAGATC CAAAGTCAAT TCATGGTCTA AACTTAATG 540
ATTTTTTTAG GTTTTGGKAC ATTTCACTGT AACTGTAGT AATTTATATC TTATTTTCCC 600
ACTAATTTAG AAAAAATATY AAATGATCCT TAATGGCAA TGGGTCTTAA GAATTTTGT 660
45 TTAAATCCCT GTTACCCAAA AGAGCCCTTT TTGTATCTC GCAGTAGTTA CAAGGATCTT 720
TCTAAATCTT AAAAAAAAAA AAAAAAGAAA GAAAGAAAAG AAAAGAAAAA AAGTCAGCCG 780
50 GGCGTGGTGG CTCATGCTG TAATCCAGC ACTTTGGGAC CAAGGTGGAC AGATCAGCAG 840
GTCAGGAGAT GGAGACCATC CCGGCAACA TGGAGAAACC CTGTCTCTAC TAAAAAAAAA 900
AAAAACTCGA GGGGGGCCCC GTACCCAATN CGCGGCTAG TGGTCGTAAA ACAATCAAA 959

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1446 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

5	CGGGGCAGGG CTGTGTGGCA CCGCCAGGGA GCGGGCCAC CTGAGTCACT TTATTGGGTT	60
10	CAGTCAACAC TTCTTGCTC CCTGTTTCT CTCTGTGGG ATGATCTCAG ATGCAGGGC	120
	TGGTTTGGG GTTTTCCTGC TTGTGCCAAG GCCTGGACAC TGCTGGGGG CTGGAAGCC	180
15	CCTCCCTTCC TGTCTTCTG TGGCTCCAT CCCCTCATGG GTGCTGCCAT CCTTCCTGGA	240
	GAGAGGGAGG TGAAAGCTGG TGTGAGCCCA GTGGGTTCCT GCCACTCAC CCAGGAGCTG	300
20	GCTGGGCCAG GACCGGAGA GGGAGCACTG CTGCCCTCCT GGCCCTGCTC CTTCGCCAGT	360
	TAGGGGTGGA CCGAGCCTCG CTTCCTCCAC TGTCTGGAG GGAAGGGGAA GGAGGGGTC	420
	TTCAGGCTGG AGCCAGGCTG GGGGTGCTGG GTGGAGAGAT GAGATTAGG GGGTGCCTCA	480
25	TGGGGTGGG AGGCCTGGG TGAAATRAGA AAGGCCAGA ACGTGCAGGT CTGCGGAGG	540
	GAAGTGTCTT GAGTGAAGGA GGGGACCCC ATCCTGGGG ATGCTGGGAG TGAGTGAGTG	600
30	AGATGGCTGA GTGAGGTTA TGGGAGCCT GAGGTTTAT GGCCTGTGT ATCCCTTCT	660
	CCCGGCCCCA GCCTGCCTCC CTCCTGCCG CCTGGCCAC AGGTCTCCT CTGGTCCCTG	720
	TCCCTCTGTT GGTGGGGAT GGAGCGGCAG CAAGGGTGT AATGGGCTG GGTCTGTCT	780
35	TCTACAGGCC ACCCCGAGGT CCTCAGTGT TGCTGGGA GCCGACGGG GCTCCTGAGG	840
	GGTACAGGTT GGGTGGCCC TCCCTGAGG TCTGGGTCA GGCTTTGGCT CTGCTGCTC	900
40	TCAGTACCA AGTCACCTCC CTCGAAAAT CCAGTCCCTT CTTTGATGT CCTTGTGAGT	960
	CACTCTGGGC CTGGCTGTG TCCCTCCTCA GCTTCTGTT CCTGGACAA GGGTCAAGCC	1020
	AGGATGGGCC CAGGCCTGG ATCCCCACC CCAGGACCC CAGGCCCTT CCCCTGCTG	1080
45	TTTGGGGG GCAGGGCAGA AATGGACTCC TTTGGGTCC CCGAGTGGG GTCCCTTCC	1140
	AGCCCTGCAT CCTCGTGCC STAGACCTG TCCCAGAGG AGGGCCTTG ACCACAGGA	1200
50	CGTGTGGTG CGCTGGCAC TCAGGACCC CCAGCTGCC CAGCCCTGGT CTCTGGGCA	1260
	TCTCTTCCCT CTGTCCGA AGATCTGCG CTCTAGTGC TTTGAGGGG TTCCCATCAT	1320
	CCCTCCCTGA TATGTATTG AAAATATTAT GCACACTGTT CATGCTTCTA CTAATCAATA	1380
55	AACGCTTTAT TTAAAGCCAA AAAAAAAAAA AAAAAACTG AGGGGGGGC CGTACCCAAT	1440
	TCGCCA	1446

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 1471 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

10 CAAAAAATAA TAATGATAAT TTAAAAATAA TAAGTAACTA ATAAAAAGAT TTTATATCCC 60
AGTCTTATGA TGTGGTGG CAAGGCTAGA TAAAAAGATG TTAGAATGAA AGAACATATT 120
15 TTTAGTGATA TGTAAATGAA GGATTCTACA ATAGTCATAT ATTTTATAT GAATGAATGT 180
TGGGTGGGC TGGAGAGGTA TGTGTGTGTA AATATAAAGG TCTCATTTC AGAGTATAGC 240
20 TCTGAAATAA TGGAACTCAT GTCTACAATT CAACATGCAT CTGTATAGTT ACATCTCATG 300
TAAATATACA CAGACATATT TTGCAGCCAG TAATTGACAG TTAATGTCCA AAACAGGTGA 360
TTGATAGSTA ACAGAAATTA GATAACCACC AATTTGCCC AAGAGAAAGA CTAGAAGGAC 420
25 TAAAAGCAGT TGAATGTATG GFACTGACAT TGTCTAAGC AGTCTGATAA CCAGTTTATT 480
GAAACGTGTG CATTAAACAGA GAATTTAATT TTAAACCAT AATTTCTCCT ATCCATTAAA 540
30 ATATTATAAT TGTTAGTAGT ATGAAACCAA CAGGAAATGT TTTTAAATCA TTTAGTGAGG 600
TGATTCAITTT GTTTCATGGG CAAACACTAT CCAGGAAAAG CCTGCTTGC CTGTTTCCCA 660
AAGAGCTCTA AGAAATAGAA TCAAGTGTA AATGGTTCAG ACCATTTCAGG ATTTCTTGTC 720
35 ACTCTTCTCA ACCCGATCT TCTGTTATT ACTGATGTT GAAACCCTGT CATTAGCCCC 780
GGCCTGGTTA AAGCCCCTCA GAGTCACCTC TCATTCATAG CAATAGAATT CAACCCCAAG 840
40 TGGTTGATGG TGTCCCAGC ACAGCCGAGA GACCTGATCT CTGGATTGAG TGCTTTTAGC 900
TCTTCGAGTT TACCCTAAGA TACCTTCGGG CAATATTTT AACCAACCCA AAAGCTCTTC 960
AGGTCATTTC TGAAGAGGAC AAGGTGAATC TTGGCTTGA ACACCATTTT TGGGCTCTTG 1020
45 CTACTGAATG AATCAGAAAG GAATTTTTC TGAAGAGCAT TAGAAAGTAA AGGAGATGTT 1080
AAAATAAGTT CTGAAGTAT GTTTTATATT TATCTAAAAC ACTGATTTTA AAAGTTTACA 1140
50 TTCAAATGTG TATTCAAAG AAGTACTGAT TTGTAATTAT TATAGTTTGT GTGTATCATC 1200
CCCTTTTAAC CGTGCCTAAC AACTGTACTT AAATTTTGTT TTCCTAGTGT AACAAATGTT 1260
TCCATAAGA TTTTCTAGAG CCAAATAATG GGAGTGAAA ATTCTTAAG TGTATATAA 1320
55 GAAATATAT TAGAAATCA GCTTTGGATT ATACGATTTT TAAATATAC TAATACAGAA 1380
TCCTCAGTAA TATGTTTGA ATTGGATTTT TTCTCAGAAC TGTTACATAA TAAATAATAC 1440
60 ATCAACCAGA AAAAAAAAAA AAAAAAATTN C 1471

5 (2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1402 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

15 AGGGACGTCT TGCCTGAGGA GATGCCCAIT TCTGTCTGG RTTACCCTCA CTGCGTGGTG 60
CATGAGCTGC CAGAGCTGAC GCGGAGAGT TTGGAAGCAG GTGACAGTAA CCAATTTTGC 120
TGGAGGAACC TCTTTTCTTG TATCAATCTG CTTGGGATCT TGAACAAGCT GACAAAGTGG 180
20 AAGCATTCAA GGACAATGAT GCTGGTGGTG TTCAAGTCAG CCCCATCTT GAAGCGGGCC 240
CTAAAGGTGA AACAAAGCCAT GATGCAGCTC TATGTGCTGA AGCTGCTCAA GGTACAGACC 300
25 AAATACTTGG GCGGCAGTG GCGAAAGAGC AACATGAAGA CCATGTCTGC CATCTACCAG 360
AAGGTGCGGC ATCGGCTGAA CGACGACTGG GCATACGGCA ATGATCTTGA TGCCCGGCTT 420
TGGGACTTCC AGGCAGAGGA GTGTGCCCTT CGTGCCAACA TTGAACGCTT CAACGCCCGG 480
30 CGCTATGACC GGGCCACAG CAACCCTGAC TTCTGCCAG TGGACAACTG CCTGCAGAGT 540
GTCTGGGCC AACGGGTGGA CTTCCCTGAG GACTTTCAGA TGAACATGA CCTCTGGTTA 600
35 GAAAGGGAGG TCTTCTCCAA GCCCATTTCC TGGGAAGAGC TGCTGCAGTG AGGCTGTITG 660
TTAGGGGACT GAAATGGAGA GAAAAGATGA TCTGAAGGTA CCTGTGGGAC TGTCTTAGTT 720
CATGTCTGCA GTGCTCCCAT CCCCCACCAG GTGGCAGCAC AGCCCCACTG TGTCTTCCGC 780
40 AGTCTGTCTT GGGCTTGGGT GAGCCAGCT TGACCTCCCC TTGGTTCCCA GGTCTCTGCT 840
CCGAAGCAGT CATCTCTGCC TGAGATCCAT TCTTCTTTTA MTCCCCCAM CCTCTCTCT 900
45 TGGATATGGT TGGTTTGGC TCATTTTACA ATCAGCCCAA GGYTGGGAAA GCTGGAATGG 960
GATGGGAACC CTTCCGCGGT GCATCTRAAT TTCAGGGGTC ATGCTGATGC CTCTGAGAG 1020
ATACAAATCC TTGCCCTTGT CAGCTTGCAA AGGAGGAGAG TTTAGGATTA GGGCCAGGGC 1080
50 CAGAAAGTCG GTATCTTGGT TGTGCTCTGG GGTGGGGTG GGTGTCTTCT GATGTTATTC 1140
CAGCCTCTG CTACATTATA TCCAGAAGTA ATTGCGGAGG CTCCTTCAGC TGCCTCAGCA 1200
55 CTTTGATTTT GGACAGGGAC AAGGTAGGAA GAGAAGCTTC CCTTAACCAG AGGGGCCATT 1260
TTTCTTTTG GCTTTCGAGG GCCTGTAAAT ATCTATATAT AATCTGTGT GTATTCTGTG 1320
TCATGTTGGG GTTTTAAATG TGATTGTGTA TTCGTTTTAC ATTAAAAAGA AGCAAAAATA 1380
60

ATAAAAAAAAAAAAAAAAAA CT

1402

5

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 1047 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

15

GGCACAGGGG ACTACAGGCA CCCACGACCA TACCCAGCTA ATTTTGTAT TTTTGTAG 60

AGATGGGGTT TCACGATGTC GCCCAGGCTG GTCTTGAAGT CCTGGGCTTG AGCGATCTTC 120

20

CCATCTTTCC ATCTTGGCCT CCTAAAGTGC TGGGACTGCA GGCATGAGCC ACCATGCCCA 180

GCCAAGATTG TTATTGATTA CCATGTTGCT TCAAGAAGCC AAGCCAGTTT CCAATATTCC 240

25

CCATTGCTG GAGTCTTGGT ACTTTGGGTA GAAGCAACTG GTAAATGTGT AATTGGAACA 300

NTTGGTGGTG TAGATAACCA CGTATGGCCA AACCTAGAGC ATCTAGGCTC ACAATTACTA 360

TCTGACTTG ATAACAAGTG TTCTGATATT AACCTGAAAA TGGGAATAAT GCCAAATCTG 420

30

TGTAACTTAA CATCTATATA CACAGTGGGG AGAAGTGAAG TTATTAAACC TGGAACTCTT 480

GTGATCAAGG CTAACAGTAG TTATCTAAGA AGCAAAGGAC CTACAATTCT TAGACTTGGA 540

GTCATATTCT TTAAGGACGT GTTCTGAAAC TATATCAAGC ATCTGGTTTC CACGTATTTC 600

35

TCCCTCAGAA ATTATGAAGT ACAAGTAAAA ATGAAGGTAC AGGGTAAGAC ACATGCTGCT 660

TTCTTGCTCT TGAGTGGAGA CAGTTTTCCT GCCATCTTAA CCCCTTWACA CAAAACAATT 720

40

TGTGTTTAT AGCAAATAAG TGAATCAACA TAATTTCAAT ATGATGTTTA TCCACCAGTA 780

CTTTCCTTTC AGCTTCTAGT CCCATAARIG GTTGTGAAG TCATCGGTTA CATTAGCCAA 840

GATAGGCCCTA GACTTGAAGT CTAGAATGTT TTTCCACTA TATGCCAAAG TAGAATGTGG 900

45

GTATCTCAGG GTCATTTTTC TTGTTCAATT TCCCACCTGT ACAGTTGTTA TGATTCATT 960

TCCTTATGTC TCTAATAAAT CTGTTCCAT GAAATGATCA AAAAAAAAAA AAAAAAACT 1020

50

CGAGGGGGGG CCCGGTACCC AAATCGC 1047

55

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 990 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

5 TTGGAAAGGG TCTAGCTCTT TCTCATTAC CAACTATATT AGAAGCACIT GAGGGAAATT 60
TACCACTCCA AATCCAAAGC AATGAACAGT CTTTCTGGA TGATTTTATT GCCTGTGTCC 120
10 CAGGATCAAG TGGTGAAGG CTTGCAAGGT GGCTTCAGCC AGATTCATAT GCGGATCCTC 180
AGAAAACATC TTTGATCCTG GAATAAGGAT GATATTGTT GTGGTTGGCC TACCACCATA 240
ACTGTTCAAA CAAAGACCA GTATGGGGAT GTGGTACATG TTCCAATAT GAAGGTAATT 300
15 ATAAGTGGAT TAAATTAGCA GACATCTATA TACTGGCTGC AATGACTGAT AAAATTTTAG 360
AAATGCCAAG TGCTGAGRGT CCATTTGTTC TACCTCTTT ATATAAAGG TGATGCTGAA 420
AGTTTGTTTA AATGACTTGT TTATATTAAT TAGTCCCCAA GTGTCCAAGT TACACCTGTT 480
20 TTTTGTGTA GTTTGTCTT TACATTTGC TACCTGTAC GGGACTCAA AGGAGGGATA 540
AGAAAGTATC CATCTAAAGA GTGCTAGACA CATACAGTGA AGCCCTCAA TATGTATGA 600
25 TTGAATAAAT GCATGAAAGA ATACATTTT AAATTTTGTG TATAGTTTG AAAGACTCAA 660
GTACGTTCTG TGTTTGGTAT TACTGAAACC ACATTTTAAA AATAACACTC ATTAAGTTAG 720
AAATATATGA GTTTAGATTG TAAAGAATG AGGAATTGAA ATAGTTGTAT ACCATATTGA 780
30 TGAATATAGA GTTTTAGGA TACCTCTTAC CTGAAATATT AATAATAATG TTTNCAGAGC 840
ATATTATACA TAATTATTG TGATTTAATC TGTTAATATG AATATCTCAT TAAAACTTT 900
35 TATTTCTGAA AAAATTATAT TGAATAAAT TTTATATAGG CAGTCCCAG CCCTTCCTC 960
CTTCAAAGTT GTCTTATAGA GTGATTGGTT 990

40

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1208 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

50

55

60

TAATCGCTAC TATAGGGAAA GCTGGTCGCT GCAGGTACCG GTCOGGAATT CCGGGTCGAC 60
CCACGCGTCC GAGCGAAATG GGGCTCCGG CCCCAGGCC GGCCTCCGGC GGCTCCGGGG 120
AGGTAGACGA GCTGTTTCGAC GTAAAGAACG CTTCTACAT CGGCAGCTAC CAGCAGTGCA 180
TAAACGAGGC GCASGGGTGA AGCTRTCAAG CCCAGAGAGA GACGTGGAGA GGGACGTCTT 240
CCTGTATAGA GCGTACCTGG CGCAGAGGAA GTTCGGTGTG GTCCTGGATG AGATCAAGCC 300

CTCCTCGGCC CCTGAGCTCC AGGCCGTGCG CATGTTTGCT GACTACCTCG CCCACGAGAG 360
 TCGGAGGGAC AGCATCGTGG CCGAGCTGGA CCGAGAGATG AGCAGGAGCK TGGACGTGAC 420
 5 CAACACCACC TTCTGTCTCA TGGCCGCTC CATCTATCTC CACGACCAGA ACCCGGATGC 480
 CGCCCTGCGT GCGCTGCACC AGGGGGACAG CCTGGAGTGC ACAGCCATGA CAGTGCAGAT 540
 10 CCTGCTGAAG CTGGACCGCC TGGACCTCGC CCGGAAGGAG CTGAAGAGAA TGCAGGACCT 600
 GGACGAGGAT GCCACCCTCA CCCAGCTCGC CACTGCCTGG GTCAGCCTGG CCACGGGTGG 660
 TGAGAAGCTG CAGGATGCCT ACTACATCTT CCAGGAGATG GCTGACAAGT GCTCGCCAC 720
 15 CCTGCTGCTG CTCAATGGGC AGGCGGCTG CCACATGGCC CAGGGCGCT GGGAGGCGC 780
 TGAGGGCTG CTGCAGGAGG CGCTAGACAA GGATAGTGGC TACCCRGAGA CGCTGGTCAA 840
 20 CCTCATCGTC CTGTCCAGC ACCTKGGCAA GCCCCCTGAG GTGACAAACC GATACCTGTC 900
 CCAGCTGAAG GATGCCACA GGTCCATCC CTTCAITCAAG GAGTACCAGG CCAAGGAGAA 960
 CGACTTTGAC AGGCTGGTGC TACAGTACGC TCCAGCGCT GAGGCTGGCC CAGAGCTGTC 1020
 25 AGGACCATGA AGCCAGGACA GAGGCCAGGA GCCAGCCCTG CAGCCCTCCC CACCGGCAT 1080
 CCACCTGCAT CCTCTGTTGG CAGGAGCCCA CCCCCAGCAC CCCCATCTGT TAATAATAT 1140
 30 CTCAACTCCA RGGTGTCCA CCTGAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1200
 AAAAAAA 1208

35

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
 40 (A) LENGTH: 1922 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

GTGCTGCGCT ACTGAGCAGC GCCATGGAGG ACTCTGAAGC ACTGGGCTTC GAACACATGG 60
 GCCTCGATCC CCGGCTCCTT CAGGCTGTCA CCGATCTGGG CTGGTCGGA CCTACGCTGA 120
 50 TCCAGGAGAA GGCCATCCCA CTGGCCCTAG AAGGGAAGGA CCTCCTGGCT CGGGCCCGCA 180
 CGGGCTCCGG GAAGACGGCC GCTTATGCTA TTCCGATGCT GCAGCTGTTG CTCCATAGGA 240
 55 AGGCGACAGG TCCGTTGGTA GAACAGGCAG TGAGAGGCCT TGTCTTGTGTT CCTACCAAGG 300
 AGCTGGCAGC GCAAGCACAG TCCATGATTC AGCAGCTGGC TACCTACTGT GTCGGGATG 360
 TCCGAGTGGC CAATGTCTCA GCTGCTGAAG ACTCAGTCTC TCAGAGAGCT GTGCTGATGG 420
 60

AGAAGCCAGA TGTGGTAGTA GGGACCCCAT CTCGCATATT AAGCCACTTG CAGCAAGACA 480
 GCCTGAAACT TCGTACTCC CTGGAGCTTT TGGTGGTGGA CGAAGCTGAC CTTCTTTTTT 540
 5 CCTTTGGCTT TGAAGAAGAG CTCAAGAGTC TCCTCTGTCA CTGCCCCGG ATTTACCAGG 600
 CTTTCTCAT GTCAGCTACT TTTAACGAGG ACGTACAAGC ACTCAAGGAG CTGATATTAC 660
 10 ATAACCCGGT TACCCTTAAG TTACAGGAGT CCCAGCTGCC TGGGCCAGAC CAGTTACAGC 720
 AGTTTCAGGT GGTCTGTGAG ACTGAGGAAG ACAAATTCCT CCTGCTGTAT GCCCTGCTCA 780
 AGCTGTCAAT GATTGCGGGC AAGTCTCTGC TCTTTGTCAA CACTCTAGAA CGGAGTTACC 840
 15 GGCTACGCCT GTTCTTGAA CAGTTCAGCA TCCCCACCTG TGTGCTCAAT GGAGAGCTTC 900
 CACTGCGCTC CAGGTGCCAC ATCATCTCAC AGTTCAACCA AGGCTTCTAC GACTGTGTCA 960
 TAGCAACTGA TGCTGAAGTC CTGGGGGCCC CAGTCAAGGG CAAGCGTCGG GGCCGAGGGC 1020
 20 CNAAGGGGA CAAGGCCTCT GATCCGGAAG CAGGTGTGGC CCGGGGCATA GACTTCCACC 1080
 ATGTGTCTGC TGTGCTCAAC TTTGATCTTC CCCCACCCC TGAGGCCTAC ATCCATCGAG 1140
 25 CTGGCAGGAC AGCAGCGCT AACAACCCAG GCATAGTCTT AACCTTTGTG CTTCCACGG 1200
 AGCAGTTCCA CTTAGGCAAG ATTGAGGAGC TTCTCAGTGG AGAGAACAGG GGCCCCATTC 1260
 TGCTCCCCA CCAGTTCGGG ATGGAGGAGA TCGAGGGCTT CCGCTATCGC TGCAGGGATG 1320
 30 CCATGCGCTC AGTGACTAAG CAGGCCATTC GGGAGGCAAG ATTGAAGGAG ATCAAGGAAG 1380
 AGCTTCTGCA TTCTGAGAAG CTTAAGACAT ACTTTGAAGA CAACCCTAGG GACCTCCAGC 1440
 35 TGCTGCGGCA TGACCTACCT TTGCACCCCG CAGTGGTGAA GCCCCACCTG GGCCATGTC 1500
 CTGACTACCT GGTTCCTCCT GCTCTCCGTG GCCTGGTTCG CCCTCACAAG AAGCGGAAGA 1560
 AGCTGTCTTC CTCTGTAGG AAGGCCAAGA GAGCAAAGTC CCAGAACCCA CTGCGCAGCT 1620
 40 TCAAGCACAA AGGAAAGAAA TTCAGACCCA CAGCCAAGCC CTCCTGAGGT TGTGGGCCT 1680
 CTCGAGGCT GAGCACATTG TGGAGCACAG GCTTACACCC TTCGTGGACA GGCGAGGCTC 1740
 45 TGGTGTCTAC TGCACAGCCT GAACAGACAG TTCTGGGGCC GGCAGTGCTG GGCCCTTTAG 1800
 CTCCTTGGA CTTCCAAGCT GGCATCTTGC CCCTTGACAA CAGAATAAAA ATTTAGCTG 1860
 50 CCCCAGAAAA AAAAAAAAAA AAAAAAATC GAGGGGGGGC CCGTACCCAA TTCGCCCTAT 1920
 AA 1922

55

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1951 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

5	TCGTCCCCAG AGCGGGCTGA GCCCCAGGCG SAGGGTGGCG GGGGAGCCTG GGGGAGCCGC	60
	CGCCACCTCC ACGGGCCTCT CTGAGCTCGG ACACCAGCGC CCTGTCCTAT GACTCTGTCA	120
10	AGTACACGCT GGTGGTAGAT GAGCATGCAC AGCTGGAGCT GGTGAGCCTG CGCGTGCTT	180
	CGGAGACTAC AGTGACGAGA GTGACTCTGC CACCGTCTAT GACAACTGTG CCTCCGTCTC	240
15	CTCGCCCTAT GAGTCGGCCA TGGAGAGGA ATATGAGGAG GCCCCGCGGC CCCAGCCCCC	300
	TGCTTGCTC TCCGAGGAAC TCCAGCCTG ATGAACCCGA CGTCCATTTC TCCAAGAAAT	360
	TCCTGAACGT YTTTCATGAGT GGCCGCTCCC GCTCCTCCAG TGCTGAGTCC TTCGGGCTGT	420
20	TCTCCTGCAT CATCAACGGG GAGGAGCAGG AGCAGACCCA CCGGGCCATA TTCAGGTTTG	480
	TGCCTCGACA CGAAGACGAA CTTGAGCTGG AAGTGGATGA CCCTCTGCTA GTGGAGCTCC	540
	AGGCTGAAGA CTA CTGCTGAC GAGGCTTACA ACATGCGCAC TGGTGCCCGG GGTGTCTTTC	600
25	CTGCCTATTA CGCCATCGAG GTCACCAAGG AGCCCCGAGCA CATGGCAGCC CTGGCCAAAA	660
	ACAGTGACTG GGTGGACCAG TTCCGGGTGA AGTTCTCTGGG CTCAGTCCAG GTTCCCTATC	720
30	ACAAGGGCAA TGACGTCTC TGTCCTGCTA TGCAAAGAT TGCCACCACC CGCCGGCTCA	780
	CCGTGCACCT TAACCCGCCC TCCAGCTGTG TCCTGGAGAT CAGCGTGCGG GGTGTGAAGA	840
	TAGGGGTCAA GGCCGATGAC TCCAGGAGG CCAAGGGGAA TAAATGTAGC CACTTTTTC	900
35	AGTTAAAAA CATCTCTTTC TCGGATATC ATCCAAAGAA CAACAAGTAC TTTGGGTTCA	960
	TCACCAAGCA CCCCCTGAC CACCGGTTTG CTGCCACGT CTTTGTGTCT GAAGACTCCA	1020
40	CCAAAGCCCT GGCAGAGTCC GTGGGGAGAG CATTCAGCA GTTCTACAAG CAGTTTGTGG	1080
	AGTACACCTG CCCCACAGAA GATATCTACC TGGAGTAGCT GTGCAGCCCC GCCCTCTGCG	1140
45	TCCCCAGCC CTCAGGCCAG TGCCAGGACA GCTGGCTGCT GACAGGATGT GGCAGTCTT	1200
	GAGGAGGGG ACCTGCCACC GCCAGAGGAC AAGGAAGTGG GCGCTGGCC CAGGGTAGGG	1260
	GAGGGTGGG CAATGGGGAG AGGCAATGC AGTTTATTGT AATATATGGG ATTAGATTCA	1320
50	TCTATGGAGG GCAGAGTGGG CTGCCTGGG ATTGGGAGG ACAGGGCTTG GGGAGCAGGT	1380
	CTCTGGCAGA GAAGGATGTC CGTCCAGGA GCACACGGCC CTGCCCCATC CTGGGCCTTA	1440
	CCTCCCCTGC CAGGGCTCGG GCGCTGTGGC TCCTGCCTTG ATGAAGCCCG TGTCTGCCT	1500
55	TGATGAAGCC TGTGCCACCT GCAAGTGCCC GCCCTGCCCC TGCCCCAACC CCCACCGAAG	1560
	AGCCCTGAGC TCAGGCTGAG CCCAGCCACC TCCCAAGGAC TTTCAGTGA GGAAATGGCA	1620
60	ACACGTGGAG GTGAAGTCCC TGTCTCAGC TCCGTATCT GCGGGGCTTC TGGTGGCTC	1680

CTGCCACTGA CCTCACCGGC ATGCTGGCCT GTGGCAGGCC TAGGACCTCA GGCGGGGAGG 1740
 AGGAGCTGCC GCAAGGCCCT GTCCCAGCAG AAGAGGGAGG CTTCTGACT GACACAGGCC 1800
 5 AGCCCCATCT TGGTCTCTGTC ACCCTGGCCC CAACTATTAA AGTGCCATTT CCTGTCAAAA 1860
 AAAAAAAAA AAAATCGGG GGGGCCCGGA ANCCAATTTC CCCAAAAAG GGGGGTTATA 1920
 10 AAAATTCN GGCNGTGTTC TAAAAATTC G 1951

15 (2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3989 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

25 GGCACAGGCC GCAGGNAACC TATGGGCGCA TATAGTTGT AATGAACTG TAGTCTCAGT 60
 TGAAGCCTA GACATGAAAT GGGTCAGTGA GCAAGGCTCT ATTCTAGTC TCCAGCCATG 120
 CCTGTGGAAC CTGARCCRC TCTCAGCACA TTGGACCCAG GCAGATGYAA AAAATTCACA 180
 30 GAACTATGAT TTGACTCAA GGGTTGTAG ATTTCTCTCT TCATTCTAAT TTCAGTGTCT 240
 AAAATCTTG CATCCRTGAA CGAGCTGGGC ATTTGATGAG ACAGGGCYGA ATACTGCAGT 300
 35 TTTCTCTCTA GAAATCATCT GGGCAITTT CTTTGAAGT ATGGGAACAA TAGGCATAA 360
 CTGTTGAC AAACCTGGGA TAATGATTT TGGGATAACG ATCTACCAGA ATGGGGATAT 420
 TTCACCTTG GTTCTGAGAT GCAAACCAA GAATATCATG ACCAGCTTTC AGGCCTCTG 480
 40 AAGTATATCT CTCACATGT CCTGTTCTCA TGCTGAGGAG CCTGAGATCC CTGTGTGGG 540
 ATTAGACAGT GGACTGTAT GGGTGTAGT GAATTGGCTT ATTTGTCTG TCCCTGTCTG 600
 45 AATGTATTGC AGGAAYTAA AAGGACCAAG AAGAGGAAGA AGACCAAGC CCACCATGCC 660
 CCAGGCTCAG CAGGGAGCTG CTGGAGGTAG TAGAGCCTGA AGTCTGTCAG GACTCACTGG 720
 ATAGATGTTA TTCAACTCT TCCAGTTGTC TTGAACAGC TGAATCTGC CAGCCCTATG 780
 50 GAAGTCTCTT TTATGCATG GAGGAAAAAC ATGTTGGCTT TTCTCTGAC GTGGGAGAAA 840
 TTGAAAAGAA GGGGAAGGG AAGAAAAGAA GGGGAAGAAG ATCAAAGAAG GAAAGAAGAA 900
 55 GGGGAAGAAA AGAAGGGGAA GAAGATCAA ACCCACCATG CCCAGGCTC AGCAGGGAGC 960
 TGCTGGATGA GAAAGRGCT GAAGTCTGTC AGGACTCACT GGATAGATGT TATCAACTC 1020
 60 CTTCACTGT GTTGAAGTGT GTGACTCATG CCAGCCCTAC AGAAGTGCTT TTTATGTATT 1080

	GGAGCAACAG CATGTTGGCT TGGCTGTTGA CATGGATGAA ATTGAAAAGT ACCAAGAAGT	1140
	GGAAGAAGAC CAAGACCCAT CATGCCCCAG GCTCAGCAGG GAGCTGCTGG ATGAGAAAGA	1200
5	GCCTGAAGTC TTGCAGGACT CACTGGATAG ATGTTATTCG ACTCCTTCAG GTTATCTTGA	1260
	ACTGCCTGAC TTAGGCCAGC CCTACAGCAG TGCKGTTTAC TCATTGGAGG AMCAKTACCT	1320
	TGGCTTKKCT CTTGACGTGG ASAAATIGAA AAGAAGGGGA AGGGGAARAA AAGAAGGGGA	1380
10	AGAAGATCAA AGAAGGAAAG AAGAAGGGGA AGAAAAGAAG GGGAAGAAGA TCAAAACCCA	1440
	CCATGCCCCA GGCTCAGCAG GGAGCTGCTG GATGAGAAAG GGCTGAAGT CTTGCAGGAC	1500
15	TCACTGGATA GATGTTATTC AACTCCTTCA GGTGTCTTTG AACTGACTGA CTCATGCCAG	1560
	CCCTACAGAA GTGCCCTTTTA YRTATTGGAG CAACAGYGTG TTGGCTTGGC TGTGACATG	1620
	GATGAAATTG AAAAGTACCA AGAAGTGGAA GAAGACCAAG ACCCATCATG CCCCAGGCTC	1680
20	AGCAGGGAGC TGCTGGATGA GAAAGAGCCT GAAGTCTTGC AGGACTCACT GGATAGATGT	1740
	TATTGACTTC CTTCAGGTTA TCTTGAAC TGCTGACTTAG GCCAGCCCTA CAGCAGTGCT	1800
25	GTTTACTCAT TGGAGGAACA GTACCTTGGC TTGGCTCTTG ACGTGGACAG AATTAAAAAG	1860
	GACCAAGAAG AGGAAGAAGA CCAAGGCCCA CCATGCCCCA GGCTCAGCAG GGAGCTGCTG	1920
	GAGGTAGTAG AGCCTGAAGT CTTGCAGGAC TCACTGGATA GATGTTATTC AACTCCTTCC	1980
30	AGTTGTCTTG AACAGCCTGA CTCCTGCCAG CCTATGGAA GTTCCTTTTA TGCATTGGAG	2040
	GAAAAACATG TTGGCTTTTC TCTTGACGTG GGAGAAATTG AAAAGAAGGG GAAGGGGAAG	2100
35	AAAAGAAGGG GAAGAAGATC AAMGAAGRAA AGAAGAAGGG GAAGAAAAGA AGGGGAAGAA	2160
	GATCAAAACC CACCATGCCC CAGGCTCAAC GGCCTGCTGA TGGAAGTGA AGAGCSTGAA	2220
	GTCTTACAGG ACTCACTGGA TAGATGTTAT TCGACTCCGT CAATGTACTT TGAACCTT	2280
40	GACTCATTCC AGCACTACAG AAGTGTTT TACTCATTG AGGAACAGCA CATCAGCTTC	2340
	GCCCTTACG TGGACAATAG GTTTTTTACT TTGACGGTGA CAAGTCTCCA CCTGGTGTTC	2400
45	CAGATGGGAG TCATATTCCC ACAATAAGCA GCCCTTASTA AKCCGAGAGA TGTATTCTCT	2460
	GCAGGCAGGA CCIATAGGCA MGTGAAGATT TGAATGAAAG TACAGTTCCA TTTGGAAGCC	2520
	CAGACATAGG ATGGGTCACT GGCATGGCT CTATTCTAT TCTCAAACCA TGCCAGTGGC	2580
50	AACCTGTGCT CAGTCTGAAG ACAATGGACC CACGTTAGGT GTGACACGTT CACATAACTG	2640
	TGCAGCACAT GCCGGGAGTG ATCAGTCRGA CATTTTAAIT TGAACCAGT ATCTCTGGGT	2700
55	AGCTACAAAA TTCTCAGGG ATTTCAITTT GCAGGCATGT CTCTGAGCTT CTATACCTGC	2760
	TCAAGGTCAK TGTATCTTT GTGTTAGCT CATCCAAAGG TGTACCTTG GTTTCAATGA	2820
60	ACCTAACCTC ATTCTTTGTG TCTTCAGTG TGGCTTGTG TAGCTGATCC ATCTGTAACA	2880

	CAGGAGGGAT CCTGGCTGA GGATTGTATT TCAGAACCAC CAACTGCTCT TGACAATTGT	2940
	TAACCCGCTA GRCTCCTTTG GTTAGAGAAG CCACAGTCCT TCAGCCTCCA ATTGGTGTCA	3000
5	GTACTTAGGA AGACCACAGC TAGATGGACA AACAGCATTG GGAGGCCTTA GCCCTGCTCC	3060
	TCTCRAITCC ATCCTGTAGA GAACAGGAGT CAGGAGCCGC TGGCAGGAGA CAGCATGTCA	3120
	CCCAGGACTC TGCCGGTGCA GAATATGAAC AAYGCCATGT TCTTGCAGAA AACGCTTAGC	3180
10	CTGAGTTTCA TAGGAGGTAA TCACCAGACA ACTGCAGAAT GTRGARCACT GAGCAGGACA	3240
	GCTGACCTGT CTCCTTCACA TAGTCCATRT CACCACAAAT CACACAACAA AAAGGAGARG	3300
15	AGATATTTTG GGTTCAAAAA AAGTAAAAAG ATAATGTAGC TGCATTTCTT TAGTTATTTT	3360
	GARCCCCAAA TATTTCTCTA TCTTTTGTGT GTTGTCTATG ATGGTGGTGA CATGGACTTG	3420
	TTTATAGAGG ACAGGTCAGC TGTCTGGCTC AGTGATCTAC ATTCTGAAGT TGTCTGAAAA	3480
20	TGTCTTCATG ATTAAATICA GCCTAAACGT TTTGCCGGGA AACTGCAGA GACAATGCTG	3540
	TGAGTTTCCA ACCTYAGCCC ATCTGCGGGC AGAGAAGGTC TAGTTTGTCC ATCASCATTA	3600
25	TCATGATATC AGGACTGGTT ACTTGGTTAA GGAGGGTCT AGGAGATCTG TCCCTTTTAG	3660
	AGACACCTTA CTTATAATGA AGTATTGGG AGGGTGGTTT TCAAAATTAG AAATGTCTG	3720
	TATTCCRATG ATCATCCTGT AAACATTTTA TCATTTATTA ATCATCCCTG CCTGTGTCTA	3780
30	TTATTATATT CATATCTCTA CGCTGGAAC TTTCTGCTC AATGTTTACT GTGCCTTTGT	3840
	TTTGTCTAGT GTGTGTTGTT GAAAAAAAAA ACATTCTCTG CCTGAGTTT AATTTTGTG	3900
35	CAAAGTTATT TTAATCTATA CAATTAAAAG CTTTGCCTA TCAAAAAAAAAA AAAAAAAAAA	3960
	AAAAAAAAAA AAAAAGCGGA CGCGTGGGC	3989

40

(2) INFORMATION FOR SEQ ID NO: 29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3735 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

	CTGCTGTTTCG CTGGCTGGGC TCCGCAGCAG GCTTGGCCAG CSGCTGACGG GTCGGCGGGC	60
	GGGTTTGTGT GAACAGGCAC GCAGCTGCAG ATTTTATTTCT GGTAGTGCAN CCTCTCAAA	120
55	GGTTGAAGGA ACTGATGTAA CAGGGATTGA AGAAGTAGTA ATTCCAAAAA AGAAACTTG	180
	GGATAAAGTA GCCGTTCTTC AGGCACTTGC ATCCACAGTA AACAGGGATA CCACAGCTGT	240
60	GCCTTATGTG TTCAAGATG ATCCTTACCT TATGCCAGCA TCATCTTTGG AATCTCGTTC	300

	ATTTTACTG GCAAAGAAAT CCGGGAGAA TGTGGCCAAG TTTATTATTA ATTCATACCC	360
	CAAATATTTT CAGAAGGACA TAGCTGAACC TCATATACCG TGTTTAATGC CTGAGTACTT	420
5	TGAACCTCAG ATCAAAGACA TAAGTGAAGC CGCCTGAAG GAACGAATTG AGCTCAGAAA	480
	AGTCAAAGCC TCTGTGGACA TGTGTGATCA GCTTTTGCAA GCAGGAACCA CTGTGTCTCT	540
10	TGAAACAACA AATAGTCTCT TGGATTIWTG GTGTACTAT GGTGACCAGG AGCCTCAAC	600
	TGATTACCAT TTTCAACAAA CTGGACAGTC AGAAGCATTG GAAGAGGAAA ATGATGAGAC	660
	ATCTAGGAGG AAAGCTGGTC ATCAGTTTGG AGTTACATGG CGAGCAAAA ACAACGCTGA	720
15	GAGAATCTTT TCTCTAATGC CAGAGAAAA TGAACATTCC TATTGCACAA TGATCCGAGG	780
	AATGGTGAAG CACCGAGCTT ATGAGCAGGC ATTAACTTG TACACTGAGT TACTAAACAA	840
20	CAGACTCCAT GCTGATGTAT ACACATTTAA TGCATTGATT GAAGCAACAG TATGTGCGAT	900
	AAATGAGAAA TTTGAGGAAA AATGGAGTAA AATACTGGAG CTGCTAAGAC ACATGGTTGC	960
	ACAGAAGGTG AAACCAATC TTCAGACTTT TAATACCATT CTGAAATGTC TCCGAAGATT	1020
25	TCATGTGTTT GCAAGATCGC CAGCCTTACA GGTTTACGT GAAATGAAAG CCATTGGAAT	1080
	AGAACCCTCG CTTGCAACAT ATCACCATAT TATTCGCCTG TTTGATCAAC CTGGAGACCC	1140
30	TTTAAAGAGA TCATCCTTCA TCATTTATGA TATAATGAAT GAATTAATGG GAAAGAGATT	1200
	TTCTCCAAAG GACCCGATG ATGATAAGTT TTTTCAGTCA GCCATGAGCA TATGCTCATC	1260
	TCTCAGAGAT CTAGAACTTG CCTACCAAGT ACATGGCCTT TTA AAAACCG GAGACAACTG	1320
35	GAAATTCATT GGACCTGATC AACATCGTAA TTTCTATTAT TCCAAGTTCT TCGATTIGAT	1380
	TTGTCTAATG GAACAAATTG ATGTTACCTT GAAGTGGTAT GAGGACCTGA TACCTTCAGC	1440
40	CTACTTTCCC CACTCCCAA CAATGATACA TCTTCTCAA GCATTGGATG TGGCCAATG	1500
	GCTAGAAGTG ATTCTAAAA TTTGAAAGA TAGTAAAGAA TATGGTCATA CTTTCCGCAG	1560
	TGACCTGAGA GAAGAGATCC TGATGCTCAT GGCAAGGGAC AAGCACCCAC CAGAGCTTCA	1620
45	GGTGGCAFTT GCTGACTGTG CTGCTGATAT CAAATCTGCG TATGAAAGCC AACCCATCAG	1680
	ACAGACTGCT CAGGATTGGC CAGCCACCTC TCTCACTGT ATAGCTATCC TCTTTTAAAG	1740
50	GGCTGGGAGA ACTCAGGAAG CCTGGAAAAT GTTGGGGCTT TTCAGGAAGC ATAATAAGAT	1800
	TCCTAGAAGT GAGTTGCTGA ATGAGCTTAT GGACAGTGCA AAAGTGTCTA ACAGCCCTTC	1860
	CCAGGCCATT GAAGTAGTAG AGCTGGCAAG TGCTTCAGC TTACCTATTT GTGAGGGCCT	1920
55	CACCCAGAGA GTAATGAGTG ATTTTCCAAT CAACCAGGAA CAAAAGGAAG CCCTAAGTAA	1980
	TCTAACTGCA TTGACCAAGT ACAGTGATAC TGACAGCAGC AGTGACAGCG ACAGTGACAC	2040
60	CAGTGAAGGC AAATGAAAGT GGAGATTTCAG GAGCAGCAAT GGTCTCACCA TAGCTGCTGG	2100

	AATCACACCT GAGAAGCTGAG ATATACCAAT ATTTAACATT GTTACAAAGA AGAAAAGATA	2160
	CAGATTTGGT GAATTTGTTA CTGTGAGGTA CAGTCAGTAC ACAGCTGACT TATGTAGATT	2220
5	TAAGCTGCTA ATATGCTACT TAACCATCTA TTAATGCACC ATTAAAGGCT TAGCATTTAA	2280
	GTAGCAACAT TCGGTTTTC AGACACATGG TGAGGTCCAT GGCTCTTGTC ATCAGGATAA	2340
10	GCCTGCACAC CTAGAGTGTG GGTGAGCTGA CCTCACGATG CTGTCTCGT GCGATTGCCC	2400
	TCTCTGCTG CTGGACTTCT GCCTTTGTTG GCCTGATGTG CTGCTGTGAT GCTGGTCCTT	2460
	CATCTTAGGT GTTCATGCAG TTCTAACACA GTTGGGGTTG GGTCAATAGT TTCCCAATTT	2520
15	CAGGATATTT CGATGTCAGA AATAACGCAT CTTAGGAATG ACTAAACAAG ATAATGGCAG	2580
	TTTAGGCTGC ACAACTGGTA AAATGACTGT AGATAAATGT TGTAATTAGT GTACACGTTT	2640
20	GTATTTTTGT TAATATAGCC GCTGCCATAG TTTTCTAACT TGAACAGCCA TGAATGTTTC	2700
	ATGCTCCCT TTTTMTTGT TCTATAGCTG TTACCTATTT TAGTGGTTGA AATGAGAGCT	2760
	AGTGATGACA GAAGGATGTG GAATGCTTC TTGACATCAT TGTGTATTGC TGGTAATCAA	2820
25	GTTGGTAACG ACTACTTCTA GCAGCTCTTA CCACTATGAC TTAAGTGGTC CTGGAAGGCA	2880
	GTAAGTGGAG GTTTCAGCA TTCTGCCTT CATGAGGGCT TCTACCACTG ACCACTTTGC	2940
30	ACGTACCTGG CTCCCAGATT TACTTAGGTA CCCCACGAGT CGTCCACATA AGCAGCTTCA	3000
	TCTTTACCTT GCCAGAGTTG ACAATTATGG GATACTCTAG TCTACTTATA CTTGTGTTCC	3060
	CATCTGCTG CCATCCTCTG AAGGCCAGGA CCCAGTCATA CATCCTTAGA AACCAAAGTA	3120
35	TGGTTTTGT TTCTCTTGG AATGTCAGGT CTTAAGGCAT TTAATTGAGG GACAAAAAA	3180
	AAAAAAGCC GATATAGTAG CTAGCTACTT AAGCATCCAT GGGTATTGCT CCATATCAAA	3240
40	GCAGATTTGC AGGACAGAAA GAGTAAATTA GCCTTCAGTC TTGGTTTACA GCTTCCAAGG	3300
	AGAGCCTTGG CCACCTGAAA TGTTAACTCG GTCCCTTCCT GTCTCTAGTT CATCAGCACC	3360
	TGCAGATGCC TGACTCTTGT TAGCCTTACT ATTCAATACA GTCCTTAGAT TCACGGTATG	3420
45	CCTCTTCTTA TCCAGGCACC TATTCTGAAT CACCATGTTG CTCTGCAGCT AGAGTTGATA	3480
	GGAGAAAATC CATTTGGGTA GATGGCCTAT GAATTTGTAG TAGACTTTCA AAATGAGTGA	3540
50	TTTGTAGCT TGGTACTTTT AAGTTTGTGG TACAGATCCT CCAAACCCAT ACTCTGAGCA	3600
	ATTAAC TGCC TTGAACATAG AGAAAATTAA GGCCTCACAG GATGAGTCTC CATCTCTGT	3660
	AAATGCTTAT TTTATCATAG TCTTTAGCCN CTACTATGAG TAAAATGTTT TCTTCNGCCG	3720
55	GGTGTGGTGA CTCAC	3735

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1667 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

10 TAGTAATTCA TTAACTCCT CTTACATGAG TAGCGACAAT GAGTCAGATA TCGAAGATGA 60
AGACTTAAAG TTAGAGCTGC GACGACTACG AGATAACAT CTCAAAGAGA TTCAGGACCT 120
15 GCAGAGTCGC CAGAAGCATG AAATTGAATC TTTGTATACC AAAGTGGGCA AGGTGCCCCC 180
TGCTGTTATT ATTCCCCCAG CTGCTCCCTT TTCAGGGAGA AGACGACGAC CCACTAAAAG 240
CAAAGGCAGC AAATCTAGTC GAAGCAGTTC CTTGGGGAAT AAAAGCCCCC AGCTTTCAGG 300
20 TAACCTGTCT GGTGAGAGTG CAGCTTCAGT CTTGCACCCC CAGCAGACCC TCCACCCCTC 360
TGGCAACATC CCAGAGTCCG GGCAGAAATCA GCTGTACAG CCCCTTAAGC CATCTCCCTC 420
25 CAGTGACAAC CTCTATTCAG CCTTCACCAG TGATGGTGCC ATTTTCAGTAC CAAGCCTTTC 480
TGCTCCAGGT CAAGGAACCA GCAGCACAAA CACTGTGTTGG GCAACAGTGA ACAGCCAAGC 540
CGCCCAAGCT CAGCCTCCTG CCATGACGTC CAGCAGGAAG GGCACATTCA CAGATGACTT 600
30 GCACAAGTTG GTAGACAATT GGGCCCGAGA TGCCATGAAT CTCTCAGGCA GGAGAGGAAG 660
CAAAGGGCAC ATGAATTATG AGGGCCCTGG AATGGCAAGG AAGTTCTCTG CACCTGGGCA 720
35 ACTGTGCATC TCCATGACCT CGAACCTGGG TGGCTCTGCC CCCATCTCTG CAGCATCAGC 780
TACCTCTCTA GGTCACTTCA CCAAGTCTAT GTGCCCCCA CAGCAGTATG GCTTTCCAGC 840
TACCCCATTT GGCCTCAAT GGAGTGGGAC GGGTGGCCA GCACCACAGC CACTTGGCCA 900
40 GTTCCAACCT GTGGGAATG CCTCCTTGCA GAATTTCAAC ATCAGCAATT TGCAGAAATC 960
CATCAGCAAC CCCCAGGCT CCAACCTGCG GACCACTTAG ACCTAGAGAC ATTAAGTGAA 1020
45 TAGATCTGGG GGCAGGAGAT GGAATGCTGA GGGGTGGGT GGGGTGGGA AGTAGCCTAT 1080
ATACTAATA CTAGTCTGC ATTTAACTGG TTATTTCTTG CCAGAGGGA ATGTTTTTAA 1140
TACTGCATTG AGCCCTCAGA ATGGAGAGTC TCCCCGCTC CAGTTATTGG AATGGGAGAG 1200
50 GAAGGAAAGA ACAGCTTTT TGTCAGGGG CAGCTTCAGA CCATGCTTTC CTGTTTATCT 1260
ATACTCAGTA ATGAGGATGA GGGCTAGGAA AGTCTTGTTC ATAAGGAAGC TGGAGAACTC 1320
55 AATGTAAAT CAAACCATC TGTAAATTCG AGTGGGTGGA GCTCTTGCTT TTGGTACATG 1380
CCCTGAATCC CTCCTCCCT CAAGAATCCG AACCAAGGA CAAAAACCAC CTAAGGGCT 1440
CTCTCCTACC CTGCCCTCCT CCTTTTTTTT TACCCCTCTC TTTTTTATT TTTCTTGCT 1500
60

	CTTTAGAACC CAGTGAAAAA TACCAGGGTA CTGGGGTGCA ACTCTTTCTT ATGATAGGTC	1560
	ATTAGTGCTT TAAGCAAAAG ATATTAGCAG CTTTGACTGC AGCATTAGCA ATTAGGAAA	1620
5	AAAAAANWA AAAACTCGAG GGGGGGCCCG GTTACCCAAT TCGCCCT	1667
10	(2) INFORMATION FOR SEQ ID NO: 31:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1408 base pairs	
	(B) TYPE: nucleic acid	
15	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:	
20	ATTACACACC TGAGCACTGT GCCTGGCAAG ACCTGTCTTA ATAGATTAGA GAACCACTGA	60
	TAGATGGTCA GCTTTCTGTA GCAGTGAGAA CCCTACATT TCAAATGTGA TAGCACCTTT	120
	GCGGGGAAAC ATCACTTGGC ACATCTGCAT TCTTTTGTGA CACAGGGTCT CACTCTGTG	180
25	CCCAGGCTAG AGTGCAATGGC ACGATCTTAG CTCACTGCAA CCTCCACCTC CCAAGTTCAA	240
	GCGATTCTTC TGCTCAGCC TCCTGAGCAG CTGGGATCAC AGACATGGGC TACCATGCC	300
30	AGCTAATTTT TTGTAATTTT TGTRGTGTTG TTTTGTGTTK TAAGTAGAGA CGGGCTTTCA	360
	CCACGTTGGS CAGGCAGGTC TCGAACTCCT GAMCTCAGGT GATCCACCCA CATCTGCGTT	420
	CCAATATCTT TCTCAACATA ATGATAGCCG TAATTAATAT TTTCCAGTAC ATTTTATGTC	480
35	CTTTACACAC GAGAGTGTGA GACAGACACA AACCCAGATC TGTCTGACTC CAAAGCCCGT	540
	TTGTCAATCAT TCCTTTTACG GTATCTTATA GTGGTATCCT TTACAGAAAG ACAGCTTTTA	600
40	CCCAACAAAG ACTTAACTTC CCAGGATGCC AGAAGGACAA AGCGGGATTG CTTTAAAGRA	660
	GRAAGTTATC AAGAMCTTAT TTTATAAATG AGATTAGATA GGGAAAGGCA ATTTATCTTT	720
	ATTAAAAACT GAAAAGGCCA GCATAGGGAA GGAGTCTCTT CGGTGGTCTT TTTCAAGGAA	780
45	ATACTTCAGT TGCTTTTATT AGAAACAGAT AGTACCTAAG GTTTTGAGGT AGGWACAGCT	840
	TAAGGCATGC TAATGKTCAT GGGTCCTTCC ATAGTCATTT TKGTAATTTG GTTWACATTT	900
50	GAGCAATAGG CAGCCCTTCA CTGCTGCTGG AYTCAITCCT GCCAYTATTA CAGGTGACAG	960
	AGGAGACAGG AGGTATGTCT TTTCTATTTT TAWACATGCT TTATATTTAA CACAAGCTCT	1020
	TGGGTATCTT AGATAACAG AAGTTGCCA GCACTCCTTT TAGTGCAITG AACCTTTTAA	1080
55	CATTTAAGCA AAATAATAAA CAGTCTTTTG AGGTTCTCTA ACAATGAAAC GTGTTGAGT	1140
	GCGAGCAGCG GAATCCATGC YTCCTCTCCT GGAGTGTGCA AKAGTCCGTG GTCTGAGTA	1200
60	TCTCACACAG ATGTGGCATT TTATGTGTGA TGCTCTAATT AAGGCCATTG GTACAGAACC	1260

5 AGATTGAGAC GTCCCTCCAG AATAATGCA TTCTTTTGCA AAGGTGAATA TTTTCTCTT 1320
AAAAAATATG TATAAGGTGG TAGTTCATT TATTAGTCTT GCTAAAAAA AAAAAAAAAA 1380
ACTINGAGGG GGGGTCGGT ACCCAAT 1408

10

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 2031 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

20

AGGATATGCA TGATTCTTAA CCAGGCTATA TGTTAAAAA AAATTGGAAA ATGCAATACA 60
TTTTTTACTA TCAAACTAC AGAATGAGTA TGCAAGTTT ATTTATCAAA ATGTAATGGA 120
25 TTTTAAAGG CTGAGAAATT TTCTTATAC CTACCTTTC AGTTATTTTA ATTATACCAA 180
ATTATCAACT AGAATAGCTT CATCCATATG AAATATAAAA TGAAGAGACA CCTAGGCTCT 240
ATCAGGCTTA GGTTCCTTG AACTTATTC CACTTAATT TCTCAGTGA AGTTAAGAGG 300
30 GGTGAGAAA CAAAGAGGG GAAACTGA CAACTAACAA AACCAGCACC ACATCGCTAG 360
GTGGTGCTA CTAATTACCT TCTCAGGATT TTCTCAGAT TGAAAAGCTT ATGAGGATTT 420
35 CTGGGATTC TTAATACCT GCTGTAGT ACAGAGCTT CCTGATGATA TTTACTCTTG 480
AGCATGTG GTTGTAAAC CTAACCTTC TTTCTCCAG AGGGTGGTGA TAGAAACAGA 540
TGGTAGTAT TATGAATGA TGTCTCGTG AAATGTTGAG GGTGGGAGA AAAGACTTTA 600
40 AGGAGGAGA GCCATCTATT TGTTCCTAA AGCCACCTCT CAGCAGAATC GTCATGTTTT 660
TCTGATGCAC CGCTCTCTT CATGCCAAG ATGACTTGG AGGCAATCTC AGGAGCTGTG 720
45 GACTTAAACR TTGCAAGCA CACTGTCTT CTCAGCGTC TCTGCAAGTC AGTAGGTGTT 780
AGTAGGTTG CAAAGTCAC TGTCTCAGCA AAGTTGAACT GGGCTACCTC TCTACAGCTG 840
TTCTCTCAGA GGGAAAAATC TTGAGACCAG ATGGTGGAGC TCTGGAGTCA GAGGAAATGG 900
50 GTGTCTTCAG CACAAAGCTG CTGCTTTTAC TTCAGCCACT TCTGACATTT TTACATACCG 960
AGCCTGAGAT TRGTGATTA TCTCAAATCA AATCACTTTG ATGGAGATAA ATAATCAAAA 1020
55 CTGTTTATA GTCATTGATT TGGTGAGAAC AGTAATGGAA AATGGTGTG AAGGACTTCT 1080
CATTTTGGG GCTTTCCTC CAGAGTCTG GCTGATTGG GTTCGCTGTT CATCTGAGCC 1140
CCCAAAAGCA TTATTACTGA TACTTGCACA CAGTCAAAG CGCAGACTGG ATGGATGGTC 1200
60

TTTTATAAGG CATTTAAGGG TACACTACTG TGTTCCTG ACCATACATT TTTCTTAGCC 1260
 CCTCAAGTAA TATAGCACAG AGTTATGAAT GACAATTCCC CTAACCATTC CTCTTCATAT 1320
 5 CTGCCTCTTC CCCTTACCAT CGTAATTCTC CAAACTGGTC ATAAAGGCAC TCTGTGAAGA 1380
 TATTGGGGAC TGACATCTTA AGCTCTCACC TGGCTGCAGT AGGAAAGGCC AACTGACGA 1440
 CAAAAAATA ATTCTTTATA AAGATGATAT GGAACATGT ATCTTTGCCC TGGGTCTGGG 1500
 10 TGGGTCCAGT CAGTCTCAGA TTTACAAGCA TTTAGGAGCC TAGGTAAAAG CTGCTAGTAT 1560
 TCTTTTAAAA GTTACATTTA TGA CTGCAA TGATAGAAAA CTCCTTCCAA TTAAATGGCA 1620
 15 TTTTATAATA TTATGTGTGT ACTTCACAGT GTTAAAAATA CCCTCATACG TTATTGCATT 1680
 TGATCTTCAC AGAAAGTGCA TTTTAACCAG TACTCTGGGT GCAATAAATA ATATGTAGAA 1740
 ATTTAAGTCC TCCAATTCCA GCATATCCAG TGAGTTTGA CAGTGTGTTT ATGTGGAATG 1800
 20 TTTAAGGATA TACAATTGTA CTTTATATAA ATTGGTCTT GTTCTCTTA AATGTGACAT 1860
 GAAATAATTG TGCTGCTACA TTACTCTGGA AATTAACAGG GGAAAAGGGA AGAGCTCTTG 1920
 25 GCTCCCTTGA GGTCTGCTA GTGGTGTAG GAGTGGTTAC AACTGAGCTT TTAGTAACCA 1980
 TTTAACCGTA TGTAAGCTTG GTTCTAATT AAAAAAAT TTCTTTTCC A 2031

30

(2) INFORMATION FOR SEQ ID NO: 33:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 971 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CGCGTCGGAA CTCGGCCCGG GGACATCCAC GGGGCGCGAG TGACACGCGG GAGGGAGAGC 60
 AGTGTTCCTGC TGGAGCCGAT GCCAAAAACC ATGCATTTCT TATTCAGATT CATTGTTTTC 120
 45 TTTTATCTGT GGGGCCCTTT TACTGCTCAG AGACAAAAGA AAGAGGAGAG CACCGAAGAA 180
 GTGAAAATAG AAGTTTTCGA TGTCCAGAA AACTGCTCTA AGACAAGCAA GAAGGGAGAC 240
 50 CTACTAAATG CCCATTATGA CGGCTACCTG GCTAAAGACG GCTCGAAATT CTACTGCAGC 300
 CGGACACAAA ATGAAGGCCA CCCCAATGG TTTGTTCTTG GTGTTGGGCA AGTCATAAAA 360
 GGCCTAGACA TTGCTATGAC AGATATGTGC CCTGGAGAAA AGCGAAAAGT AGTTATACCC 420
 55 CCTTCATTTC CATACGGAAA GGAAGGCTAT GCAGAAGGCA AGATTCCACC GGATGCTACA 480
 TTGATTTTTC AGATTGAACT TTATGCTGTG ACCAAAGGAC CACGGAGCAT TGAGACATTT 540
 60 AAACAAATAG ACATGGACAA TGACAGGCAG CTCTCTAAAG CCGAGATAAA CCTCTACTTG 600

CAAAGGGAAT TTGAAAAAGA TGAGAAGCCA CGTGACAAGT CATATCAGGA TGCAGTTTAA 660
 GAAGATATTT TTAAGAAGAA TGACCATGAT GGTGATGGCT TCATTCTCC CAAGGAATAC 720
 5 AATGTATACC AACACGATGA ACTATAGCAT ATTTGTATTT CTACTTTTTT TTTTAGCTA 780
 TTTACTGTAC TTTATGTATA AAACAAAGTC ACTTTCTCC AAGTTGTATT TGCTATTTTT 840
 10 CCCCTATGAG AAGATATTTT GATCTCCCA ATACATTGAT TTTGGTATAA TAAATGTGAG 900
 GCTGTTTTGC AAACTTAAAA AAAAAMWAAA AAACTSGAG GGGGGCCCGT ACCCAANTCG 960
 CCGNATATGA T 971
 15

(2) INFORMATION FOR SEQ ID NO: 34:
 20

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1792 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:
 GAACCCCTT TCTCCTGGTA AAGGTAAGG GGGGGGATAA TGTTTACCAC AGGTACGAAA 60
 30 TAGTCACTTT AACATTGAGA CCTCTGCCTC ATTGAATTCA GGTTTTTTAA GTACTTGAAA 120
 CTCTTCAGAT TCTCCTTATT TTAGTTCTT TTTACATTTA TGAAGTAGAA AGCATTTGTTT 180
 35 TGTAAACTGT TTTGAAAATA AATAGCCTAG TCTCTTATCC TCTTTAGCGT GGATTAAAGG 240
 TGAAGTTCTG CAAATGGGAG AGTGTTCACA GTAGATAGCT CAGATTGATT GAACACATTT 300
 GAGGAAGAGA CTCCTGCATG AGATACCAGC ATTTTACAA ATACTTTTTA TGTACATTCT 360
 40 TTAITTTGTC ATTTTGTCAA CCTCTCCCC AAGCACATCT TCTTCCTTT TACTATGICT 420
 ATGTAGGGAA AAACAAAACA AAAAATGCA CTTACGTTAC ACTCCCAAAA TGTGGTAAT 480
 45 CCGTGTCTTT CAAAAACAT TTCTGTTTTT TGTTTGTTT TGGTCAGTCC ATTGCATAAG 540
 TGACAAGTTT GGTGCTTGT GGCACGTATG TATGAAGCGG GAGGGGATG ASAATTGCCT 600
 GTCCTTCAGT ARGCTGTAAA AGTAATTAC ATGTAAGTAA AAAGGGAAAA TAGAATAGAT 660
 50 GCCAAAGTCA TTTATTCAGT CCTAGTTTT CTTATGTGGC ATTACTGCAT CTGCTAGTTA 720
 GTGAGAAAGC ACCCTCAGCT TTTACTGCTC CCTCCCTGC CTGCCAACAC ACTTGATGTG 780
 55 TGCAAACAGC CCTCAAGTAT CTGTCAGATG ACCTATATAA GGTATTGAAT AAGGTATTCT 840
 TGTCAGTTTA GAAATGGACT GGATAAACT TACTTGGTTG TCATTATTTT ATCTCATTG 900
 TCCTGTTACA TGCCCTATGT TAAGATAATT ATATTGCCAC TAATAATCAA GATGCTAAAT 960
 60

GAGTATTACA ACTGGCTAAT ATCATTTTTT ATATACAAGG GTATGTGTAT ATTTGGAATT 1020
 GREATGAGAA ACTCATTTGT ACCCATTTGA GTGATATTGC ACAACAAACA CAGATAYCTA 1080
 5 CAGACTCCGT TTTCAATTTT TCGTGTCTT TATGATAATG ATCTTTGTAG ATTGGTTATT 1140
 TCTGTACTTT ATCTGTAATA AACTTTGTAG ATCCTGTGAA CCATTACTTT GCCTAAATCA 1200
 CTTGAGACTT GAGTCTTTAA TAACAAAGCA TCAATATTCA CTAAAGTCAA TCTCTTTTGA 1260
 10 GTTCTGTGA CTTGGCTAGA AGCTCTTGAC ACTAAGGGAT TAGTGTAAAT TTCCCTGGG 1320
 GGTGTTCCAC TAGGGCATT CTGTATAATG ACTTGATGTT GCCACATAGA CTTCAAGATA 1380
 15 TATAATATTT TGAGGATTTT GTTGATTGGC CTATGTTTTA TTGCATAGTG TGAAACGTGT 1440
 AAAGCTTGGT TAACCTGTAT ATAGATAGCT TATTGTTGAC TAGTTATAGT GTATTTAGGG 1500
 TTGCCTGTAA TATTTAAGCT TCTTTACTGA TGTGTGTGCT GGTAGGAACA TATAATTTTT 1560
 20 GTACATTATA TTTACTGAGA TGTGCCCTT TTTATTTTAC AAATACTTTG GAATTCGAAT 1620
 GTGTTTTTTG CTTCCGTGAG GATTAATTTG GAAAGGTTTT TAATGACATT CCACTGATTT 1680
 25 CAGATTTTGC TTGAGATTGA CTTCAATAAA TTGCTCTGTA TGTTCACAAA AAAAATTAAA 1740
 AAACGCGAGG GGGGCCCGGT ACCCAANNCG CCGGATATGA TCGTAAACAA TC 1792

30

(2) INFORMATION FOR SEQ ID NO: 35:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 896 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

AGTTGNANAC AACAGGACCT GAGTCCTTGG GCAGCACCAG TAGGTTGCCC CYTGCTCYT 60
 GCCAGCYTCA CYTGCCACYT TTTGCCCTY TCGGGATGCC TTCGCAGACA GAGYTTYTCG 120
 45 CTGCTGTGG TGGCCAYTCT TTGCTTTTGG TTYTCTTGCC CCTTGGCCTC CCTTTTGTGTC 180
 CCCGGGCAGC CTTGTGTGAC CTGCCCTTTT CCTCCCTTC CTTCCAGGA CAAGCACGCC 240
 50 GAGGAGGTGC GGAAAAACAA GGAGCTGAAG GAAGAGGCCT CCAGGTAAAG CCTAGAGGCC 300
 AAAGAACTTT CCAGGTCAGC CGGACAGCTC CAGCAGCTCC ACGTTCAGG CAGCCTCGMC 360
 CGCCGGCTGC GCTCCAGCA CTGGGTTTG GGGGAGGGG GTGGCCAAG GGGGTTTCC 420
 55 TCTGCTTTTG GTGTTGTAC ATGTTAAGAA TTGACCAAGT AAGCCATCCT ATTTGTTTCC 480
 GGGGAACAAT GACGGGTGG GARAGGGGAG AGGAGAGAGT TTGGGAAAGG GAGATGGAGA 540
 60 AGAACTCAAG GACATTGCAA CCTGCCCGG CGCAGATCTG ATTTTCACAT CTCTACCTGG 600

ACATTGAGCC TCCCAGGCAC CATGTTGAGG AGAGATGAAA ACCAGGGCGG TAGAACTTCA 660
 GGGTGAAGGA CAGAGTCCTG GGTGGGGCAG CGGCTGCAGG GCGCACCAGA GAACCCAGCC 720
 5 AGAGGGGGTG TGAGTACCAG TGGTGTTGCT TCCACCTGC AGCAGGTGGG ATGAGGTCTG 780
 TGTGTGTGTG TGAACCATCA TTTTGTGATC ATCATGACCA ATGAAACATT GAAAAAAAAA 840
 10 AAAAAAACTG GAGGGGGGCC CGTACCCAAN TCGCCGNATA GTGATCGTAA ACAATC 896

15 (2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 912 base pairs

(B) TYPE: nucleic acid

20 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

25 TCGACCCACG CGTCCGGTCA GCCAGTGCAT TCCAGCCATG ACAGCCTTCT GCTCCCTGCT 60
 CCTGCAAGCG CAGAGCCTCC TACCCAGGAC CATGGCAGCC CCCCAGGACA GCCTCAGACC 120
 AGGGGAGGAA GACGAAGGGA TGCAGTGTCT ACAGACAAAG GACTCCATGG CCAAGGGAGC 180
 30 TAGGCCCGGG GCCAKCCGCG GCAGGGCTCG CTGGGGTCTG GCCTACACGC TGCTGCACAA 240
 CCCAACCTTG CAGGTCTTCC GCAAGACGGC OCTGTTGGGT GCCAATGGTG CCCAGCCTG 300
 35 ARGGCAGGGA AKGTCAACCC ACCTGCCCAT CTGTGCTGAG GCATGTTCTT GCCTACCATC 360
 CTCCTCCCTC CCGGCTCTC CTCACAGCAT CACACCAGCC ATGCAGCCAG CAGGTCTCTC 420
 GGATCACYGT GGTTKGGTGG AGGTCTGTCT GCACTGGGAG CCTCARGARG GCTCTGCTCC 480
 40 ACCCACTTGG CTATGGGAGA GCCAGCAGGG GTTCTGGAGA AAAAACTGG TGGGTTAGGG 540
 CCTTGGTCCA GGAGCCAGTT GAGCCAGGGC AGCCACATCC AGGCGTCTCC CTACCCCTGC 600
 45 TCTGCCATCA GCCTTGAAGG GCCTCGATGA AGCCTTCTCT GGAACCACTC CAGCCCAGCT 660
 CCACCTCAGC CTTGGCCTTC ACGCTGTGGA AGCAGCCAAG GCACTTCCTC ACCCCYTCAG 720
 CGCCACGGAC CTYTYTGGGG AGTGGCCGGA AAGCTCCCSG GCCTYTG GCC TGCAGGGCAG 780
 50 CCCAAGTCAT GACTCAGACC AGGTCCACA CTGAGCTGCC CACTCTGAG AGCCAGATAT 840
 TTTTGTAGTT TTTATKCTT TGGCTATTAT GAAAGAGGTT AGTGTGTTC CTGCAATAAA 900
 55 CTGTCTCTG AG 912

60 (2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1382 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

10	AAATTCGGCAC GAGCGGAGGC GAGGGAAACT RAGGGCGAAA GTTGTGTGTC GTGTTGGCAG	60
	GAGGGCCTAG AAGGGAAAGA CTGTCTAGTG GGACAATGTC ATATTATAAA TTTGGAATGC	120
	TGAATAGAAA ATTATAGATT TTGATATTGA AGGAAATGAA GCGAAGCYTA AATGAAAATT	180
15	CAGCTCGAAG TACAGCAGGC TGTTTGCTG TTCCGTTGTT CAATCAGAAA AAGAGGAACA	240
	GACAGCCATT AACTTCTAAT CCACTTAAAG ATGATTTCAGG TATCAGTACC CCTTCTGACA	300
20	ATTATGATTT TCCTCTCTA CCTACAGATT GGGCCTGGGA AGCTGTGAAT CCAGAGTTKG	360
	CTCCTGTAAT GAAAACAGTG GACACCGGGC AAATACCACA TTCAGTTTCT CGTCCTCTGA	420
	GAAGTCAAGA TTCTGTCTTT AACTCTATTC AATCAAATAC TGGAAGAAGC CAGGGTGGTT	480
25	GGAGCTACAG AGATGGTAAC AAAAATACCA GCTTGAAAAC TTGGRATAAA AATGATTTTA	540
	AGCCTCAATG TAAACGAACA AACTTAGTGG CAAATGATGG AAAAAATTCT TGTCCAATGA	600
30	GTTCGGGAGC TCAACAACAA AAACAATTAA GAACACCTGA ACCTCCTAAC TTATCTCGCA	660
	ACAAAGAAAC CGAGCTACTC AGACAAACAC ATTCATCAAA AATATCTGGC TGCACAATGA	720
	GAGGGCTAGA CAAAACAGT GCACTACAGA CACTTAAGCC CAATTTTCAA CAAAATCAAT	780
35	ATAAGANACA AATGTTGGAT GATATTCAG AAGACAACAC CCTGAAGGAA ACCTCATTGT	840
	ATCAGTTACA GTTTAAGGAA AAAGCTAGTT CTTTAAGAAT TATTTCTGCA GTTATTGAAA	900
40	GCAATGAAGTA TTGGCGTGAA CATGCACAGA AACTGTACT TCTTTTGTGA GTATTAGCTG	960
	TTCTTGATTC AGCTGTTACA CTGGCCCAT ATTATTCGAA GACTTTTCTT ATGAGGGATG	1020
	GGAAAAATAC TCTGCCTTGT GTCTTTTATG AAATCGATCG TGAACCTTCG AGACTGATTA	1080
45	GAGGCCGAGT TCATAGATGT GTTGGCAACT ATGACCAGAA AAAGAACATT TTCCAATGTG	1140
	TTTCTGTCAG ACCGGCGTCT GTTCTGAGC AAAAACTTT CCAGGCATTT GTCAAAATTG	1200
50	CAGATGTTGA GATGCAGTAT TATATTAATG TGATGAATGA AACTTAAGTA GTGATAAAG	1260
	GAAGTTTAGC ATAAATTATA GCAGTTTCT GTTATGCTT AATTTACCAT CTCCATAGTT	1320
55	TTATAGCTAC TATTGTATTT CACTTGTGA ATTAAAGTAT TTGAATCTT TAAAAA	1380
	AA	1382

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 872 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

10 GGGCTACTTC AAAGCCCTGG GCCTTATTTTTC TTCAGGTAAA AAAATATATAA GTCAGATCTC 60
ATCCCGGCTG GCCATGCTGT TAGACCCCTTT CATCCTTCTC TTCTGCCTCT TCTCAACAGC 120
15 TGCCCGAGTCC TGTTTGAAT TCATATACAT ACAGTTCTAA TACTGATGTA TTTACCCCTCA 180
TAAGCCACTC AACCCAGAAT CTTATTTGAA TTATAATCCA GAAACATCAG GTGACGTGTG 240
AGACTACTGT ATGAGAAAGA GACAGTTTAA GGGTCAGTCC AATGGAAAAA AGAGTTCTCA 300
20 GAGCTTTCTT TAGCTTATTC TCATCAAAGA GCTTTCTCTG CAGAAGGAAC CTACTGGTTC 360
CTCCTTTCCA GTCCTAGAAA TCCTGACCTA GAGTGGCTTA ATCCTGCTAG CACCTCTCTC 420
25 TGGCACTCTG GTGCCAAATG ACTCCAGGAA CTGGGCCATG ATGTGGTGGG AATGACCTTA 480
CCCTGAGCAT GTCATCATG CATGAACAA CAGCTAAGAG CAGAGCTTAG AGCTTAGAGC 540
TGGGCCCTGT AAGGTGAGAG GAATCACATC CTGCAGAAGT CTGTCCTGAG AAGCAGGTAC 600
30 TCCTGTCACA GCAGAGACAC AGTGGATACC TGAGTAACAA TAATACAAGA CAGGACGTGG 660
GMACAGCAAA AGATTGGGT GTCAGAAGAR GCGAGAACA CTTTCAGGCA GGAACATTCA 720
35 RARTTGTCTT TGGAGGAART AGGCMCSAAG GCTGGGCAGG ATTTCTCGGG GCAGAGATGG 780
AGCAAGCAAT TGAATGAAA GCCATGGCAT GGGAAAAGGA GCACTGGCCA CAGGGAGTGC 840
AACGTTGTGA TGCAAGGCCA CTGTGGAGCC AT 872
40

(2) INFORMATION FOR SEQ ID NO: 39:

- 45 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 812 base pairs
(B) TYPE: nucleic acid
50 (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

55 GGCAGAGGCT CACCCAGCA GAGATTGAGG GGAACCGTG ATGAAATTTT TAAGTATCT 60
GCTTGATGAT AATAATTTT CTCTATGTT AATGTTGGCT CCGTTTGGGT GTTTAGCTTT 120
TGAAAGGAGT ATGAAATGC GGAATGGGC TTTGGGCTT GAGGAGGTGT GATCTCTAGT 180
60 GTTTAAAAA TTAAATGCA CAAATAGAAA TAATCACCC ACATTATTGA ACCCCACTAA 240

AGCATATCCT TTTGTCCAT ATTCCTTTCC TGCTGCCCTC GTGTGTACCA TTATTACTCA 300
 GTTGTGATTT GAGCTCGTTC CACTTAAAGT CATTCATAGA TACTTTTGGC TCGTGTTKGA 360
 5 ATATTTATTTG AATTCTTATT CTGTGTTTTA CTTAATTACT TTATTATGGA ACCTTTACAC 420
 AGGTCTGGTG TACTTGTCTT TTGAAAAGTC TTATGTTGAC CACCATCACT GAGCATATAG 480
 10 CTTTTTCCTT ATTTCTTGG GATAATTACC CGAAGTGGAA ATACCGAATC AAACCTCTGT 540
 TTTCTTTCTT TGGCACTATT ATATAAATG TTTTCCAAAC AAGGCATGTT TACAATAGAC 600
 ATTTTTCAAA ATCTGGGTAT TTGTCCTATT TTGCTCTCTG TATGCAGAAT TCAGCGGGGT 660
 15 GCCAAGTCGT TTTCTGTGTG GGTGAGAGA CAGGCTGTGC AGCCCACTGT TGCATAGGAC 720
 TAACACTAC AAATCATGCT GAGACCGAGC TATTTTGTCT GCTTAGARGC TTTGCAGCCT 780
 20 TGAGTAAGTT TCGNCATCTG GAAACNTTGN AA 812

25 (2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1515 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

35 AATTCCGCAC GAGGGAAATT CAAGCACTTT TCCTAAAAGA AGGGGGAATG GATGCTGAAA 60
 CAACACGTNT CCCACAAAGG GAGCAGACAC TGGGCTTGTG AAGCTGCCCC ATACCTTCCC 120
 CACAGAACTG GGGTCCGGCC TCCTGACAT GCAGATTTC ACCCAGAAGA CAGAGAAGGA 180
 40 GCCAGTGGTC ATGGAATGGG CTGGGGTCAA AGACTGGGTG CCTGGGAGCT GAGGCAGCCA 240
 CCGTTTCAGC CTGGCCAGCC CTCTGGACCC CGAGGTTGGA CCCTACTGTG ACACACCTAC 300
 45 CATGCGGACA CTCTTCAACC TCCTCTGGCT TGCCCTGGCC TGCAGCCCTG TTCACACTAC 360
 CCTGTCAAAG TCAGATGCCA AAAAAGCCGC CTCAAAGACG CTGCTGGAGA AGAGTCAGTT 420
 TTCAGATAAG CCGGTGCAAG ACCGGGGTTT GGTGGTGACG GACCTCAAAG CTGAGAGTGT 480
 50 GGTCTTGTAG CATGCGAGCT ACTGCTCGGC AAAGGCCCGG GACAGACACT TTGCTGGGGA 540
 TGTACTGGGC TATGTCACTC CATGGAACAG CCATGGCTAC GATGTACCA AGGTCTTTGG 600
 55 GAGCAAGTTC ACACAGATCT CACCCGTCTG GCTGCAGCTG AAGAGACSTG GCCGTGAGAT 660
 GTTTGAGTTC ACGGGCTCC ACGACGTGGA CCAAGGGTGG ATGCGAGCTG TCAGGAAGCA 720
 TGCCAAGGGC CTGCACATAG TGCCCTGGCT CCTGTTTGTG GACTGGACTT ACGATGATTT 780
 60

300

COGGAACGTC TTAGACAGTG AGGATGAGAT AGAGGAGCTG AGCAAGACCG TGGTCCAGGT 840
 GGCAAAGAAC CAGCATTTTCG ATGGCTTCGT GGTGGAGGTC TGGAAACCAGC TGCTAAGCCA 900
 5 GAAGCGCGTG ACCGACCAGC TGGGCATGTT CACGCACAAG GAGTTTGAGC AGCTGGCCCC 960
 CGTGCTGGAT GGTTCAGCC TCATGACCTA CGACTACTCT ACAGCGCATC AGCCTGGCCC 1020
 TAATGCACCC CTGTCTGGG TTCGAGCCTG CGTCCAGGTC CTGGACCCGA AGTCCAAGTG 1080
 10 GCGAAGCAA ATCCTCCTGG GGCTCAACTT CTATGGTATG GACTACGCGA CCTCCAAGGA 1140
 TGCCCGTGAG CCTGTGTGTC GGGCCAGGTA CATCCAGACA CTGAAGGACC ACAGGCCCCG 1200
 15 GATGGTGTGG GACAGCCAGG YCTCAGAGCA CTTCTTCGAG TACAAGAAGA GCCGCAGTGG 1260
 GAGGCACGTC GTCTTCTACC CAACCCTGAA GTCCTGCAG GTGCGGCTGG AGCTGGCCCC 1320
 GGAGCTGGGC GTTGGGGTCT CTATCTGGGA GCTGGGCCAG GGCCTGGACT ACTTCTACGA 1380
 20 CCTGCTCTAG GTGGGCATG CGCCTCCGC GGTGGACGTG TTCTTTTCTA AGCCATGGAG 1440
 TGAGTGAGCA GGTGTGAAAT ACAGGCCTTC ACTCCGTTAA AAAAAAAAAA AAAAAAAAAA 1500
 25 AAAAAAAAAA AAAAA 1515

30 (2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 704 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

40 AAGATGGTGG CGCCAGAGC TTCGCTCTAT GCTGCTCCC TGAGAGAGGC GTTTCATCA 60
 ACCAGTTTTC CAAGGACTTC AATGAGAGGA CAAAGGACAT CAAGGAAGGC ATTCCTCTGC 120
 CTACCAAGAT TTTAGTGAAG CCTGACAGGA CATTTGAAAT TAAGATTGGA CAGCCCACTG 180
 45 TTTCTACTT CCTGAAGGCA GCAGCTGGGA TTGAAAAGGG GGCCTGGCAA ACAGGGAAAG 240
 AGGTGGCAGG CCTGGTGACC TTGAAGCATG TGTATGAGAT TGCCCGCATC AAAGCTCAGG 300
 50 ATGAGGCATT TGCCCTGCAG GATGTACCCC TGTCGCTGTT TGTCGCTCC ATCATCGGGT 360
 CTGCCCCTTC TCTGGGCATT CGCGTGGTGA AGGACCTCAG TTCAGAAGAG CTGCGACCTT 420
 TCCAGAAGGA ACGAGCCATC TTCCTGGCTG CTCAGAAGGA GGCAGATTTC GCTGCCCAAG 480
 55 AAGAAGCTGC CAAGAAGTGA CCCTTGCCCC ACCAACTCCC AGATTTCAA GGAGGTAGTT 540
 GCAAAAGCTG TGCCCAAGGG GAGGAAGGAG GTCACACCAA TATGATGATG GTTTTCATGA 600
 60 CTTTGAATGA TATATTTTTC TACATCTAGC TGTATCGAGG CATCAGGCCT GAATAACAT 660

CCTTCTTAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAA

704

5

(2) INFORMATION FOR SEQ ID NO: 42:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1094 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

GGCAGCTTTC TTACAAACCC ATCCTTCTGA AATGTTGCTT CAAATTCATC CTCTGCTCCC 60

20

CAGTCCCCTT ATTCCACACA TACTGTTACT GTTTCCTTAT CCTACTTTCT CAATTTTGGA 120

ACATAGTTGC AGTTACTGCA TTGAATACCT GTGGGTTTGC CTGTTGTCTT GTCTGTCTCT 180

GTGGTTCTTG TAATANIGGA TCCAGAGAT AAAATGGACA GTGTGNATGC ACAGTTAATT 240

25

CAGAACTAG ACCTTACTTG CTGTGTGAAA TACCAACTAA ATTCTCAGTG AACTCAGCTG 300

ANCTTTATCT CCTTTGTGTT CCCCAATTTA TAATTTTCAGT TCAGGCCCTG AAAGATGGAA 360

30

TCCCAGCTAA GAAATACAAG TTACACCTG TACTAGCAGC CCATGTGTGC ATGTTCTTTA 420

AGTGCTCTTG CAGCTATGTC ATTTATATTG ATTTCCCTGT ATTATTATAA GCAAAGCAAA 480

TTTGAGGAAA AAAACCCATA ATACCACACC TCATTTTITT CAAGTAATAG GGTGATAAGT 540

35

CTCATYCTYC ATATAATATG TTGAGTATGC AGTATATTAT GTGTTAGGCT CTGGANAGGC 600

AGAGGTTAGA TCATGTWACA GATCATATCK GATTAGGCAG ATAAACAGTA TTTTAACCTT 660

40

TTCTTTATTA TAIGTAACTT GCTTTCAGGT TTTTAAATGT TACTATTATG TCTTTAATAT 720

ATTATCTTTA TTGTACTTTT TGTATACAGA GTGATTTTCC TTTTTTAAAA AAAATTGTGT 780

CTTTAGGATG GATTCCAAAG ATGTGGAATC AGTAGGTTTA AGGAATATGG ATATTTTGGC 840

45

TGGCAAGGTG GCTCACACCT GTAATCCCAG CACTTTGGGA GGCTGAGGTG GGTGGATCAC 900

CTGAAGTCAG GAGTTCGAGA CCAGCCTGAC CAACATGGCG AAACCTGTGTT TTTACTAAAG 960

50

ACACACWAA AATTGACCAG TGGTGGTGGC ATGTGCTTGT AGTCCACTTT AGCTACTCGA 1020

GAGGCTGAGG CAGGAGAATC GCTTGAACCC GGGAGGCAGA GGTGTCAGTG AGGCAAGATG 1080

GCACCTCTAC ACTC 1094

55

(2) INFORMATION FOR SEQ ID NO: 43:

60

(i) SEQUENCE CHARACTERISTICS:

302

(A) LENGTH: 1321 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

	TGGCTTAGGC CACACCCCTT CCTTGGCTG GAACTACTGG ACAGACCCCTT TTGAGATGTG	60
10	CCTGTGGTGC TGTGGAGATG TGTGTAGTGG TCTTAGCTCT TTGTTGAGCT TGTGTGTGTG	120
	TTGTGTAGTC TTASCTGTAT GCTGAATTG GCGTGTGTG GGAGGGCTTC TTAGCTCTTT	180
15	GGTGAGATGG TATTTCATG TGTTTGTATC ASCTGAATGT TGCTGGAAAT AAAACCTTGG	240
	TTGTMAAGG CTCATTTTGT TGGGAAGTAA GTAGGGGAAA AGGTCTTTGA GGGTTCCTAG	300
	GCTCCTTTGT ACAACAGGAA AATGCCCTCA AGCCTTGCTT CCCAGCAACC TGGGGCTGGT	360
20	TCCAGTGGC TGGTCTGGC CCTTCTGGT TCTTATCTCA AGGCAGAGCT TCTGAATTC	420
	AGGCCTTCAT TCCAGAGCCC TCTTGTGGC AGGCCTTCCT TTGCTGGAGG AAGGTACACA	480
25	GGGTGAAGCT GATCTGTATC TTGGGGGATC TCCTTGGCCT GTCCACCAA GTGAGAGAAG	540
	GTACTTACTC TGTACCTCC TGTTCAGCCA GGTGCATTAA CAGACCTCCC TACAGCTGTA	600
	GGAACTACTG TCCAGAGCT GAGGCAAGGG GATTTCCTAG GTCATTGGA GAACAAGTGC	660
30	TTTAAAGTA GTTAAAGTA GTAACCTCA CTGTATTTAG TGGGGTGGAA TTCAGAAGAA	720
	ATTGAAGAC CAGATCATGG GTGGTCTGCA TGTGAATGAA CAGGAATGAG CCGGACAGCC	780
35	TGGCTGTAT TCTTTCTTC GTCGCCATT GAGCCCTTCT CTGCCCTTAC ATTTTGTGTT	840
	CTCCATCTAC CACCATCCAC CAGTCTATTT ATTAACCTAG CAAGAGGACA AGTAAAGGGC	900
	CCTCTTGGCT TGATTTTGT TCTTTCTTC TGTGGAGGAT ATACTAAGTG CGACTTTGCC	960
40	CTATCCTATT TGGAAATCCC TAACAGAAIT GAGTTTCTA TTAAGGATCC AAAAAGAAAA	1020
	ACAAATGCT AATGAAGCCA TCAGTCAAGG GTCACATGCC AATAACAAT AAATTTTCCA	1080
45	GAAGAAATGA AATCAACTA GACAAATAAA GTAGAGCTTA TGAAATGGTT CAGTAAGGAT	1140
	GAGTTGTGTG TTTTGTGTT TGTTTGTGTT TGTTTTITA AAGACGGAGT CTCGCTCTGT	1200
	CACTCAGCTT GGAGTGCAGT GGTATGATCT TGGCTCACTG TAACCTCCGC CTCCCGGGTT	1260
50	CAAGCCATTC TCCGCTCA GTCTCTGAG TAGCTGGGAT TACAGGTGCG TGCCACCATG	1320
	CCTGGCTTAT TTTGTGTTT TTAGTAGAGA CAGGGTTTCA CCATGTTGGT CGGGCTGGTC	1380
55	TCAACTCTT GACCTCTTGA TCCGCTGCC TTGGCTCCC AAAGTGATGG GATTACAGAT	1440
	GTGAGCCACC CGTCCCTAG CCAAGGATGA GATTTTAAA GTATGTTTCA GTTCTGTGTC	1500
	ATGGTTGGA GACAGAGTAG GAAGGATATG GAAAAGGTCA TGGGAAGCA GAGGTGATTC	1560
60	ATGGCTCTGT GAAATGAGG TGAATGGTTC CTTATTGTCT AGGCCACTTG TGAAGAATAT	1620

GAGTCAGTTA TTGCCAGCCT TGGAAATTAC TTCTCTAGCT TACAATGGAC CTTTGTAACT 1680
GGAAACACCC TTGTCTGCAT TCACCTTAAA ATGTCAAAC TAATTTTAT AATAAATGTT 1740
5 TATTTTCACA TTGAAAAAA AAAAATTT AAAACYCGG GGGGGCCCS GACCCCAT 1800
NGCCCTAAG GGGGGGGT T 1821

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1024 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GGGGCACAGT TGAAGAAGCG ACCGAGGAC TGGAGTCGT TAGTGAGGAT GACGCGGCAT 60
25 GGCAAGAACT GCACCGCAGG GCCGTCTACA CCTACCAGA GAAGAAGAAG GACACAGCG 120
CCTCGGGCTA TGGGACCCAG AACATTGAC TGAGCCGGA TGCCGTGAAG GACTTCGACT 180
30 GCTGTGTCT CTCCTGCAG CCTGCCACG ATCCTGTGT CACCCAGAT GGCTACCTGT 240
ATGAGCGTGA GGCCATCTG GAGTACATC TGCACCAGAA GAAGGAGATT GCCCGGCAGA 300
TGAAGGCTA CGAGAAGCAG CGGGGCACCC GCGCGAGGA GCAGAAGGAG CTTACGCGG 360
35 CGGCCTGCA GGACCATGTG CGGGCTTCC TGGAGAAGGA GTCCGCTATC GTGAGCGGC 420
CCCTCAACCC TTTCACAGCC AAGGCCCTCT CGGGACCAG CCCAGATGAT GTCCAACTG 480
GGCCAGTGT GGGTCTCCA AGTAAGGACA AGGACAAAGT GCTGCCAGC TTCTGGATCC 540
40 CGTCGCTGAC GCCGAAGCC AAGGCCACCA AGCTGGAGAA GCGTCCCGC ACGGTGACCT 600
GCCCCATGTC AGGAAGCCC CTGCGCATGT CGGACCTGAC GCGGTGCAC TTCACACCG 660
45 TAGACAGCTC CGTGGACCG GTGGGCTCA TCACCCGAG CGAGCGCTAC GTGTGTGCCG 720
TGACCCGGA CAGCTGAGC AACGCCACCC CTTGCGCTGT GCTGCGGCC TCTGGGGCTG 780
TGGTCACCT CGAATGCGTG GAGAAGCTGA TTCGGAAGGA CATGGTGGAC CCTGTGACTG 840
50 GAGACAACT CACAGACCG GACATCATCG TGCTGCAGCG GGGCGGTACC GSTTCGCGG 900
CTCCGGAGTG AAGCTGCAAG CGAGAAATC ACGGCCGTG ATGCAGGCCT GAGTGTGTG 960
55 GGGAGACCA ATAAACCGC TTGGGTGCG AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1020
AAAA 1024

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 983 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

CGACACGGCT GCGAGAAGAC GACAGAAGGG CCCGACCGCG AGCCGTCCAG GTCTCAGTGC 60
TGTGCCCCC CCAGAGCCTA GAGGATGTTT CATGGGATCC CAGCCACGCC GGGCATAGGA 120
GCCCCTGGGA ACAAGCCGGA GCTGTATGAG GAAGTGAAGT TGTACAAGAA CGCCCGGGAG 180
AGGGAGAAGT ACGACAACAT GGCAGAGCTG TTGCGGTGG TGAAGACAAT GCAAGCCCTG 240
GAGAAGGCCT ACATCAAGGA CTGTGTCTCC CCCAGCGAGT AACTGCAGC CTGCTCCCGG 300
CTCCTGGTCC AATACAAAGC TGCCTTCAGG CAGGTCCAGG GCTCAGAAAT CAGCTCTATT 360
GACGAATTCT GCGCAAGTT CCGCTGGAC TGCCCGCTGG CCATGGAGCG GATCAAGGAG 420
GACCGGCCCA TCACCATCAA GGACGACAAG GGCAACCTCA ACCGCTGCAT CGCAGACGTG 480
GTCTCGCTCT TCATCACGGT CATGGACAAG CTGCGCCTGG AGATCCGCGC CATGGATGAG 540
ATCCAGCCCG ACCTGCGAGA GCTGATGGAG ACCATGCACC GCATGAGCCA CCTCCACCC 600
GACTTTGAGG GCCGCCAGAC GGTGAGCCAG TGGCTGCAGA CCCTGAGCGG CATGTGGCG 660
TCAGATGAGC TGGACGACTC ACAGGTGCGT CAGATGCTGT TCGACCTGGA GTCAGCTAC 720
AAGCCTTCA ACCGCTTCCT GCATGCTGA GCCCGGGCA CTAGCCCTTG CACAGAAGGG 780
CAGAGTCTGA GCGATGGCT CCTGGTCCCC TGTCGCCAC ACAGGCGTG GTCATCCACA 840
CAACTACTG TCTGCAGTG CTTGTCTGGT GTCTGTCTTT GGTGTCAGAA CTTTGGGCC 900
GGGCCCCCTC CCACAATAAA GATGCTCTCC GACCTTCAA AAAAAAAAAA AAAAAAAGR 960
KSGGGCCGGT CCCANTCCC CCC 983

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2421 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

CCGGCTGATC GCTGCCGCTC CGCCAATACA ATAGAGCCAK CCACTACCAG CAGCCTGGCC 60

	CTCTTCCTCC TTCTCCAGAG AGACCAATCC AGCCGAATC GGGGTTTGCC TGAGGAGAAG	120
	GAGGAAGTGA CCATGGACAC AAGTGAAAAC AGACCTGAAA ATGATGTTCC AGAACCTCCC	180
5	ATGCCTATTG CAGACCAAGT CAGCAATGAT GACCGCCCGG AGGGCAGTGT TGAAGATGAG	240
	GAGAAGAAAG AGAGCTCGCT GCCCAAATCA TTCAAGAGGA AGATCTCCGT TGTCTCAGCT	300
	ACCAAGGGGG TGCCAGCTGG AAACAGTGAC ACAGAGGGGG GCCAGCCTGG TCGGAAACGA	360
10	CGCTGGGGAG CCAGCACAGC CACCACACAG AAGAAACCTT CCATCAGTAT CACCACTGAA	420
	TCACTAAAGA GCCTCATCCC CGACATCAAA CCCCTGGCGG GGCAGGAGGC TGTGTGGAT	480
15	CTTCATGCTG ATGACTCTCG CATCTCTGAG GATGAGACAG AGCGTAATGG CGATGATGGG	540
	ACCCATGACA AGGGGCTGAA AATATGCCGG ACAGTCACTC AGGTAGTACC TGCAGAGGGC	600
	CAGGAGAATG GGCAGAGGGA AGAAGAGGAA GAAGAGAAGG AACCTGAAGC AGAACCTCCT	660
20	GTACCTCCCC AGGTGTCAGT AGAGGTGGCC TTGCCCCAC CTGCAGAGCA TGAAGTAAAG	720
	AAAGTGACTT TAGGAGATAC CTTAACTCGA CGTTCATTA GCCAGCAGAA GTCCGGAGTT	780
25	TCCATTACCA TTGATGACCC AGTCCGAATC GCCCAGGTGC CCTCCCCACC CCGGGCAAG	840
	ATTAGCAACA TTGTCCATAT CTCCAATTTG GTCCGTCCTT TCACMTTAGG CCAGCTAAAG	900
	GAGTTGTTGG GCGGCACAGG AACCTTGGTG GAAGAGGCTT TCTGGATTGA CAAGATCAAA	960
30	TCTCTTTGCT TTGTAACGTA CTCAACAGTA GAGGAAGCTG TTGCCACCCG CACAGCTCTG	1020
	CACGGGTCA AATGGCCCA GTCCAATCCC AAATTCCTTT GTGCTGACTA TGCCGAGCAA	1080
35	GATGAGCTGG ATTATCACCG AGGCTCTTTG GTGGACCGTC CCTCTGAAAC TAAGACAGAG	1140
	GAGCAGGGAA TACCACGGCC CCTGCACCCC CCACCCACAC CCCCGTCCA GCCACCACAG	1200
40	CACCCCGGG CAGAGCAGCG GGAGCAGGAA CCGGCAGTGC GGAACAGTG GGCAGAACGG	1260
	GAACGGGAAA TGGAGCGCGG GGAGCGGACT CGATCAGAGC GTGAATGGGA TCGGGACAAA	1320
	GTTCGAGAAG GGCCCGTTC CGATCAAGG TCCCGTRACC GCCCGCGCAA GGAACGTCCG	1380
45	AAGTCTAAAG AAAAGAAGAG TGAGAAGAAA GAGAAAGCCC AGGAGGAACC ACCTGCCAAG	1440
	CTGCTGGATG ACCTTTTCCG AAAGACCAAG GCAGCTCCCT GCATCTATTG GCTCCACTG	1500
	ACTGACAGCC AGATCGTTCA GAAAGAGGCA GAGCGGGCCG AACGGGCCAA GGAGCGGGAG	1560
50	AAGCGCGGAA AGGAGCAAGA AGAAGAAGAG CAAAAGGAGC GGGAGAAGGA AGCCGAGCGG	1620
	GAACGGAACC GACAGCTGGA GCGAGAGAAA CGTCGGAGC ACAGTCGGGA GAGGACAGG	1680
55	GAGAGAGAGA GAGAAAGGGA GCGGACAGG GGGACCGAG ATCGGGATAG GGAAGGGAC	1740
	CGAGAACGAG GCAGGGAAAG GGATCGCAGG GACACCAAGC GCCACAGCAG AAGCCGAGT	1800
60	CGGAGCACAC CTGTGCGGGA CCGGGGTGGG CGCCGCTAGC TGGGAAAACA CTAGAGCTGC	1860

AGGTACCAGC CACTCGGCCC CAGGGGGTTA TGGCCACAGA GGGATAGGCA CAGTCTCCAC 1920
CACCCCTGGAG CCAAGGGTCT TTCACATCAC CTATCOCTAC ATACATACCA AATGGAAAAG 1930
5 TGGCCATCCT TTTCCCCCA AACACACCCC CTTAACCTAT CTCTTGGGAC TTAGCCCGAC 2040
CCTCCCTCTC ATTTCCCATTT AAGTCTGAGA GGCAAGAGCT AGGTTAGGCA AGGAGGTGGT 2100
TGGCCAGAGA TGGGGAACAG CCAGGTGCCC CAGTCCCTCTG ATTTTTCCTC CATCCTGCTT 2160
10 ACCAECTCCC TGGGTACTTA CAGCCTTCTC TTGGGAACAG CCGGGGCCAG GACTGGGTCA 2220
CCTATGAGCT GAATCAGCAT CTCCTCCTGA GTCCAGGGC CCCTGCAGTT CCCAGTCTCT 2280
15 TCTGTCTGC AGCCCTTGCC TCTTTCCAC AGGTTCCACT TTATATCCAC CTTTTCCTTT 2340
TGTTCAATTT TTATTTTAT TTTTITATT ATTAAATGAT GTGGTCTATG GAAAAAAAAA 2400
TAAAAATCTG ACTTAGTTTT A 2421
20

(2) INFORMATION FOR SEQ ID NO: 47:

25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 840 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

30

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

35

CTCAAACTCC TGAGCTGAAG CGATCTACCT GCCTCAGCTA GGATTACAGG TGTGAGCCAC 50
CGCACCCAAC CTCAATAAGC KTATTTGATA AAKATATGC AAGCTCCCTT TATKCACTTT 120
TCATTGAGAA TGTTTAGTAA TTGTATTGT TTTTCAGATT TTCAGCCAA TATATCTCTT 180
40 TGGCCACTGT GTCAGTGTAT TCTACCTAWA CATCATCAGG TGTTTCTGCT ATTGGCTGTA 240
TGATGGAACA CTGCGGCTCA TTTTCTGAA AACTGCCGAT AGTGCATAGA RTGCTGGGAT 300
GGAAACCAGA ARCTTTGAAT TCAAGCCTTG GTCTGCTT GTTTTGTCTT GGGTGGCCTT 360
45 GAGTCAGCCA CATACCTTTT AAAATCTCAA TTTATTAGAA ATTATTCAG AATCAAAATCA 420
AATGAGAAGG TATATACAAA AGTGCTTTAT CCCACAATA ACTATTCAG AGAGAGCAA 480
50 GGAGAGGACA TTTACTCAAC ACCTCCTAAA AGGCAGCCAG TGAAATTAGG CATTTTATTT 540
AATCCTCCTG GCAACTCTGA GAGTAAAGCA TTATTAATCC CATTTTGGCT GTTTAAAGAA 600
ATTATTTGCA CTAGATTCCA GCTGTAGTTT AGYTTGAGAA AAAAAATCC TGAGATGTGA 660
55 ATTACAGCT TTCTGGGTTT AAAGCCCAAG CTCTATCACA TCATGCTATT ATTGTTACAT 720
TACTGCTAGT TCTATGAAA GAAATACTAA TTTATGAAAT ACATCTTATC CAAAAAATA 780
60 AAAAAAATC TGGGAGGGG GGGCGTACC CAAATCGCCG GATAGTGATC GTAAACAATC 840

5 (2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2432 base pairs

(B) TYPE: nucleic acid

10 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

15 GGCACGAGGC CCGGAACGCT GAGGAAGGGC CCGTCCCGCC TTCCCGGGCG CGCCATGGAG 60
CCCCGGGCGG TTGCAGAAGC CGTGGAGACG GGTGAGGAGG ATGTGATTAT GGAAGCTCTG 120
CGGTCATACA ACCAGGAGCA CTCCCAGAGC TTCACGTTTG ATGATGCCCA ACAGGAGGAC 180
20 CGGAAGAGAC TGGCGGASTG CTGGTCTCCG TCCTGGAACA GGGCTTGCCA CCTCCCACC 240
GTGTCACTG GCTGCAGAGT GTCGAATCC TGTCCCGGA CCGCAACTGC CTGGACCCGT 300
25 TCACCAGCCG CCAGAGCCTG CAGGCAYTAG CCTGYTATGY TGACATCTCT GTCTCTGAGG 360
GGTCCGTCCT AGAGTCGCA GACATGGATG TTGTA CTGGA GTCCCTCAAG TGCCTGTGCA 420
ACCTCGTGCT CAGCAGCCCT GTGGACAGA TGCTGGCAGC AGAGGCCCGC CTAGTGGTGA 480
30 AGCTCACAGA GCGTGTGGGG CTGTACCGTG AGAGGAGCTT CCCCCACGAT GTCCAGTTCT 540
TTGACTTGGG GCTCCTCTTC CTGCTAACGG CACTCCGCAC CGATGTGCGC CANAGCTGTT 600
35 TCAGGAGCTG AAAGGAGTGC GCCTGCTAAC TGACACACTG GAGCTGACGC TGGGGGTGAC 660
TCCTGAAGGG AACCCCCCAC CCACGCTCCT TCCTTCCCAA GAGACTGAGC GGGCCATGGA 720
GATCCTCAAA GTGCTCTTCA ACATCACCTT GGACTCCATC AAGGGGGAGG TGGACGAGGA 780
40 AGACGCTGCC CTTTACCGAC ACCTGGGGAC CCTTCTCCGG CACTGTGTGA TGATCGCTAC 840
TGCTGGAGAC CGCACAGAGG AGTCCACGG CCACGCAGTA ASCCTCCTGG GGAAC TTGCC 900
45 CCTCAAGTGT CTGGATGTTT TCCTCACCTT GGAGCCACAT GGAGACTCCA CGGAGTTCAT 960
GGGAGTGAAT ATGGATGTGA TTCGTGCCCT CCTCATCTTC CTAGAGAAGC GTTTCACAA 1020
GACACACAGG CTGAAGGAGA GTGTAGCTCC CGTGCTGAGC GTGCTGACTG AATGTGCCCC 1080
50 GATGCACCGC CCAGCCAGGA AGTTCTTGAA GGCCAGGTG CTGCCCTTC TCGGGGATGT 1140
GAGGACACGG CCTGAGGTTG GGGAGATGCT GCGGAACAAG CTTGTCCGCC TCATGACACA 1200
55 CCTGGACACA GATGTGAAGA GGGTGGCTGC CGAGTTCTTG TTTGTCTGT GCTCTGAGAG 1260
TGTGCCCGA TTCATCAAGT ACACAGGCTA TGGGAATGCT GCTGGCCTTC TGGCTGCCAG 1320
GGGCCTCATG GCAGGAGGCG GCCCGAGGGC AGTACTCAGA GGATGAGGAC ACAGACACAG 1380
60

308

ATGAGTACAA GGAAGCCAAA GCCAGCATAA ACCCTGTGAC CGGGAGGGTG GAGGAGAAGC 1440
 CGCCTAACC TATGGAGGGC ATGACAGAGG AGCAGAAGGA GCACGAGGCC ATGAAGCTGG 1500
 5 TGACCATGTT TGACAAGCTC TCCAGGAACA GAGTCATCCA GCCAATGGGG ATGAGTCCCC 1560
 GGGGTCATCT TACGTCCCTG CAGGATGCCA TGTGCGAGAC TATGGAGCAG CAGCTCTCCT 1620
 CGGACCCCTGA CTCGGACCCCT GACTGAGGAT GGCAGCTCTT CTGCTCCCC ATCAGGACTG 1680
 10 GTGCTGCTTC CAGAGACTTC CTTGGGGTTG CAACCTGGGG AAGCCACATC CCACTGGATC 1740
 CACACCCGCC CCCACTTCTC CATCTTAGAA ACCCCTTCTC TTGACTCCCG TTCTGTTTAT 1800
 15 GATTTCCTC TGGTCCAGTT TCTCATCTCT GGAATGCAAC GGTCTTCTTG TGCTAGAACT 1860
 CAGGCTCAGC CTGGAATTC ACAGACGAAG TACTTTCTTT TGTCTGCGCC AAGAGGAATG 1920
 TGTTTCAAG CTGCTGCTG AGGGCAGGGC CTACCTGGGC ACACAGAAGA GCATATGGGA 1980
 20 GGGCAGGGT TTGGGTGTGG GTGCACAAA AGCAAGCACC ATCTGGGATT GGCACACTGG 2040
 CAGAGCMANT GTKTTGGGGT ATGTGCTGCA CTTCCAGGG AGAAAACCTG TCAGAACTTT 2100
 25 CCATACGAGT ATATCAGAAC ACACCTTCC AAGGTATGTA TGCTCTGTTG TTCCTGTCTT 2160
 GTCTTCACTG AGCGCAGGGC TGGAGGCCTC TTAGACATTC TCCTTGGTCC TCGTTCAGCT 2220
 GCCCACTGTA GTATCCACAG TGCCCGAGTT CTGCTGGTT TTGGCAATTA AACCTCCTTC 2280
 30 CTACTGGTTT AGACTACACT TACAACAAGG AAAATGCCCC TCGTGTGACC ATAGATTGAG 2340
 ATTTATACCA CATACCACAC ATAGCCACAG AAACATCATC TTGAAATAAA GAAGAGTTTT 2400
 35 GGACAAAAAA AAAAAAAAAA AAAAAAAAAA AA 2432

40 (2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1742 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

50 GTCTTGCAGG AGCTGCACGC GCGGAGGGT CGCANGAACA AGGAGCAGCG AGAAGAGATG 60
 TCGGGCTAAG GCGCCGSAC GGGGGCGCC CATCTGCGA CGGAACACGT TCGGGTTTGT 120
 55 GTTTTGTTC GTTCACTCT GTCTAGATGC AACTTTTGT CTCTCTCCC CACCCAGCC 180
 CCCAGCTTCA TGCTTCTCT CCGCACTCAG CCGCCCTGCC CTGTCTCTGT GGTGAGTCC 240
 TGACCAAGGC TTTCCCTGCA GGAGCCGCC GCGGTGRAGA CCGGTCCCT CCGTGCAGAC 300
 60 ACCAGGCCCG GCGGGCTGG GTCCCCCGG GCGCTGTGA GAGAGGTGGY GGTGACCGTG 360

	GTAAACCCAG GGCCTGGCG TGGGATCRCG GGTCTTACG CTGGGCTGTC TGGTCAGCAC	420
	GTGCAGGTCA GGGCAGGTCC TCTGAGCCGG CGCCCCCTGGC CAGCAGGCCA GGCTACAGTA	480
5	CCTGCTGTCT TTTCCAGGGG AAGGGGCTCC CCATGAGGRA GGGGCGACGG GGGAGGGGGG	540
	TGATGGTGCC TGGGAAGCCT GCKTGTGCAN CCGTGTCTTG TTGAACTGGC AGGCGGGTGG	600
10	GTGGGGGCTG CAGCTTTCTT TAATGTGGTT GCACAGGGGT CCTCTRAGAC CACCTGGCGT	660
	GAGGTGGACA CCTGGGCCT TCCTGGAAGC CTGCAGTTGG GGGCCTGCCC TGAGTCTGCT	720
	GGGAGTGGG CATCTCTGTC CAGGGACCCA TGAGCAGGCT GCATGGTCTA GAGGTGTGG	780
15	GCAGCATGGA CAGTCCCCCA CTCAGAAGTG CAAGAGTTCC AAAGAGCCTC TGGCCCAGGC	840
	CCCTCCGTGG GACAGCCCCG CGCCCTCTCC CCACCAGGGC TTTGCAGATG TCCTTGAAAG	900
20	ACCCACCTTA GAGCCCTTGG GAGTGTGCGC CCCTCTGTG CCCTCTGCCC TGGTGGAAAG	960
	GGCASCACAA GTCCTCTCA GGGAGCCCCA AGGGGGAATT TKTGGGACCG CTGCCACAG	1020
	ATCCAGGTGT TGGAAAGGCA GCGGTAAGG TTCCCAAGCC AGCCCCAACA CCCTTCCAC	1080
25	TTGGCACCCA GAGGGGGCTG TGGTGGAGG CCGTACTCCA GGCTCTCTCT GCCCACACCC	1140
	TCTGGGCTGA GTTCTTCTT TCCCTTGGAC GCCAGTGTG GGCCTTGGAG GACGGTCAGC	1200
30	TGGAGGATGG CCGTGGGGGA GGCTGTCTTT GTACCACTGC AGCATCCCCC ACTTCTCCAC	1260
	GGAAGCCCCA TCCCAAAGCT GCTGCCTGGC CCCTTGTGT AAAGTGTGAA GGGGCGGGCT	1320
	GAGTTCTCTT AGGACCCAGA GCCAGGGCCC TCAACTTCCA TCCTGCGGGA GGCCTTGGCC	1380
35	GGGCACTGCC AGTGTCTTCC AGAGCCACAC CCAGGGACCA CGGAGGATC CTGACCCCTG	1440
	CAGGGCTCAG GGGTCAGCAG GGACCCACTG CCCCATCTCC CTCTCCCCAC CAAGACAGCC	1500
40	CCAGAAGGAG CAGCCAGCTG GGATGGGAAC CCAAGGCTGT CCACATCTGG CTTTGTGGG	1560
	ACTCAGAAAG GGAAGCAGAA CTGAGGGCTG GGATATTCTT CATGGTGGCA GCGCTCATAG	1620
	CGAAAGCCTA CTGTAATATG CACCATCTC ATCCACGTAG TAAAGTGAAC TTAAAAATTC	1680
45	AATCAAATGA ACAATTAAAT AACACCTGT GTGTTTAAGA AAAAAAAAAA AAAAAAATG	1740
	CG	1742
50		

(2) INFORMATION FOR SEQ ID NO: 50:

- 55 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1487 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- 60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

	GGCAGAGCC TCCGCGAACT GTGGAGTCGG CGGAGGGCTG GAATCAGCGT GGGCTCCAGG	60
5	TCGCTGGCAG CCGGGTGGCA GAACTCTTCC GAGGCTCCTT GGGAAGAAGC TACACCCGAG	120
	GGAGCCCGAT GGGCCTCGAA AACCTGGCCC GCTCTGGTTC TGTACCATTG CAAGGGGAAC	180
10	CGTAAACTGA GCTTTTCTAA CGTGGGTTTC TGCCAAGTAC TTTTCCAGCT GGGCCCTTCC	240
	CCCCAGCACA CAGGAGAGCC TCTGTGTAGC CAGCGCTTGA CAGTCGTTAG GTAGGTTGTA	300
	CTGTGTAGGG AGGAGCTCAA GATCATGAAT GGTGTGCACA GGAGAAAGCG GTTGCATCTT	360
15	TGCAAACTA TATACCTGCT GTGGTTGTG TTTTCTTTTC TGCTGAGTAA TGAAGTTGTA	420
	AGTTCACACT GGCACATTCT CAGGGCTGTG CAGATTATTT GCACTTTATT TCATAGGTGR	480
20	ATAAGTGCTT TTTAGCTTTC TTTGTATATT GAGTTGCTTT TGAATTGCTT CCCATATTTT	540
	TATTTTCATAC AAAGTGAACA ATTGTGGCCC CTCTATTTTA TTTATAAAGG TTCAGTGTAT	600
	CTTTGCCCTGC CTACATCAAT CTGCAAGGGA GTTGCAGAAA GCCTCATGTT CATCGAGCCG	660
25	TGAGTCACAA CCAATTCTTA AGCTGTTATA ACAAAAAAGT GTTTGCTTTT TTTTACAAGT	720
	AACTTTAAAA GTGTAGTTTA GAAAGAAAAC ATTTTCAATA AAAAGACACT ACATTAAATCC	780
30	TGGATGCTTG CAAATCCTAA AATMTATTC TCCTCTAGCG TTGCACAGCT CTGTGTTGTA	840
	TACACAGACT AGCTTTAAAA TTTGTACAT ACCACTTTAC CTTTACTTTT ATGTATCATT	900
	CCCCGACTT CCTTACTGCA GGTGTGGGCA AGAAAACTTT TCCTTTAACA CTTTCAACA	960
35	GCGGGCATAA AATTCTGCAG CTGAGGTCTT GAAGAATGCA GATGGGTACA GTATGTGTTG	1020
	GAGCTCACAG TGTGTATTGA CTAACCTAGT TCCTTTTTTG CTTTTTGTG TATTGCTTG	1080
40	TTAAAGTGA CTCCAGGTA GCAACTCTCT TTTTAAAGG TGGGAACGAA AGGGACGTAG	1140
	GAAGAATAGA TCTAGATTAT TTAACAGTCT TCGATAGAGT TTGAAAGCTT TCTTCTTCAT	1200
	TCAATTTTGG GCAAATACT GCCTCTGCAT TTGTTTATAA CAAAAAGATT AGATTAAATA	1260
45	GTAGCTTTTG TTGGTGGAAA TTACCAGCTC TATAAGTCAC CCTTGGTGGT TCATGGACCT	1320
	CTGATTAGCT TGGGTTTTC AGTCTCATG CCACATGTAT ATGTGGAGCC AATGGCCTTT	1380
50	TGGTGCTCAG CTGTTTACGT CTGACTCCTT GACTTCTTTG GTACAGTGAT GGAGTCAGAT	1440
	CTCATTAAGT GTGATTCTCC ATGGATATAA CCAGCCCCAA AAAAANG	1487

55

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1328 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

5	GGCAGGAGCT CGTGCCGAAT TGGCAGCAG AGAAGATTG AAGAAGCCAG ATCCAGCTTC	60
	CCTGCGGGCT GCTTCTGTG GGAAGGGAA AAAGAGGAAG GCCTGTAAGA ACTGCACCTG	120
10	TGGCCTTGCC GAAGAACTGG AAAAAGAGAA GTCAAGGGAA CAGATGAGCT CCCAACCCAA	180
	GTCAGCTTGT GGAAACTGCT ACCTGGGCGA TGCCTTCCGC TGTGCCAGCT GCCCCTACCT	240
	TGGGATGCCA GCCTTCAAAC CTGGGAAAAA GGTGCTTCTG AGTGATAGCA ATCTTCATGA	300
15	TGCCTAGGAG GTTCTTGACA TGGGACCCAT CTGCTCCTCC AGCCAACCTCC TGTCCCTCAC	360
	ATCCCACCAT GGTGGCTCCT CCCACCTCCT CTGGATTGT TCACTCTGAG ATCTGTTTGC	420
20	AGAGTGGGTG CTTAGCAGAC AGAGTGAAGC TGGCTGGGGG GCACAGTGGT GTGTAGTGCT	480
	GCTGTGTATC AAAAGACCAA GGTATTATGG GACCTGGTTT CAGAATGGGA TGGGTTTCTT	540
	CACCTCATGT TAAGAGAAGG GAGTGTGTCC TGAAGAAGCC CTTCTTCTGA TGTAAAAATG	600
25	CTGACCAGAA CGCTCTTGAG CCCAGGCATC GTTGAGCATT AACACTCTGT GACAGAGCTG	660
	CAGACCCCTG CCTGAGTCT CATCTCAGCA ATGCTGCCAC CCTCTGTCTT TTCAGAGTTG	720
30	TTAGTTTACT CCATTCTTTG TGACACGAGT CAAGTGGCTC ACAACCTCCT CAGGGCACCA	780
	GAGGACTCAC TCACTGGTTG CTGTGATGAT ATCCAGTGT CCTCTGCCCC CTTCCATCCC	840
	CAACCACATT TGACTGTAGC ATGTCATCTG TGTCTGTGTG TCATTATGT TAACCTTCAG	900
35	GTATTAAACT TGCTGCATAT CTTGACATAT CTTGAGATTG TGCATGTCTT GTAAAGAGAG	960
	GGGATGTGCA TTTGTGTGTG ATGTTGGATA GTCATCCACG CTCAGTTTGG ACCATTGGAG	1020
40	GAACCTTAGTG TCACGCACAA ATGGGGCTAT TCCTACGCTT AGAATAGGGC TTGTCTGCCC	1080
	ACTTTAGAAG AGTCCCAGGT TGGTAGCAT TTAGAGGGAA GCAGGGCAGA ACTCTGAACG	1140
	ACAATACGTC TCTCTGAGCA GAGACCCCTT TGTCTTGTT ATCCACCCAT ATGGACTTGG	1200
45	AATCAATCTT GCCAATAAT TGGAGAGATT GTGTGGATT AAGAGACCTG GATTTTATA	1260
	TTTTACCAGT AAATAAAAGT TTTCATTGAT ATCTGTCTT GAAAAAAAAA AAAAAAAAAA	1320
50	AAACTCGA	1328

55 (2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1856 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

5	GAATTCGGCA CGAGCTGTGC AACATTGCAA ATGAACTGGT AGCCGAGGGT TCCGCTGCCC	60
	CCTAGATTAA ATTCCCCGGG CTGAATCTGA GTGGCAGAT TAAATATCA TATTTAAAT	120
	TGCTGTCTTC AATTAAACCA TTTATGACCA TAATTAATT TCAAGATGTC GATGCATGCT	180
10	TTCCAGGCC TTCCTCTTT GTACAAAGT AAATGTGCA AAAGCGTTTC ACTTATATC	240
	TTCAAACATG ATGCTAATTT AAATTAATTA CTTCCTATGA TAGGTATCA TTCCTATGAT	300
15	TTTGCCACTG TTATAGTTC TCTCAAAAT ACATCTAGG AAGAGGATTA TTTTAAGTA	360
	TTTGATTATC TTCTATCTC TTTATTTAA TTCTCATTA CTGAGAAAT TCGTTCCATT	420
	GGTTGGCATT GATACAGTAA ATTGTAAAT GAGGAGACA TAAATAAAT CTAAATTACT	480
20	TGTGCTTAAT GACTGTAGCA GAATSCCTTT TCTCTAATC AGTTGTCTT TCTTGCAGTT	540
	TAGTTTGATA GATTGTCAAG CTATGCTGCT TCCATGAGT TAGTTCGCT GGTAGGAACG	600
25	CAGGCTTCTT TGTCTGGGT TGTAGCTTGC ATGATGCCC CTTAGGCAG ACAACGTAGC	660
	CGGAGATCAC AAATCAGGCC CTGGGTGAG TTCTATCTT CTGAGGTGC AGAGAGGTG	720
	GCAGAACTG AACTCACTGG GCAAGGCTGG CCAAGGACT GATCTTTAA TGCACCTAT	780
30	GTGTTGAGGA AGCCACAGGC CATATTGCA TCTGAGAGG AAATAAGAG GAAAAACCC	840
	ACAAAGTATA ACACCCCTT AAGATAGTC TATTTTAAE TGAATTAAT TTTTCAGTT	900
35	ATACCATTGG CCATTACAA GATATAATG TTCAATTTCT TAAAGATCC TTTGTGACT	960
	TGTCTTTTCA TCTTTGCTA TTTATATTTG TCACTGTTAG TCAACAAAGT CTTATTGCT	1020
	GAGGAAGGAC TTTCTGACAC TTTCTGTAG ACATCAACA CTGGGAGGG TGGTGTATA	1080
40	CTTTTAAAAA AATGTTATTC TGATTAAAC AATAATATG GCTTTTTC A TGAAAGAGC	1140
	GCCACCTTGC AAGGTTTAGT GAGATTATG GAATGCAAC AACTAGCAG GAATTGCTGC	1200
45	TAGCTCCAAA AATTTCGAA GCAAAAGCTA GCGCCATTTG GTTGGAAGT TTGAACTGA	1260
	TTAACAGATT TGCAATTGAA GTGACTGAG ACATTAGGT CAGCATTAG TTAAAAATAG	1320
	AAAGAGGAAT AAGACATCT YTTCTCTTA GAAGAGATA CACCCAACT AATAATCCTT	1380
50	CCCACTTTCA TTGAGATCAG CTTGTCTGAT AACCTGATA GAGTGTGACA ATGATAAACA	1440
	TGATAATAGT GGTACTTTTG TAATTTTGT GGTGCAATTA AAGAGATAGT AAAGATGAG	1500
55	TTCACTTTT CTYGAACAT YCTATTCCT AGATGTAGT TACCTCAAT TGGGAATTAT	1560
	AACTGTCTTA ATTTTGTGTG TGTACCTGA TGCCCTTTT GCTTTAATAC CCACAGTGA	1620
60	ACAATTAAAT ATCACACTAT GACATATGAT TTAGGTAGA TTTTAAAG ATAAATTITA	1680

	GGGGTAAATG TTTACTTCAA AATGACTCCA TATTTCAAAT ATCTGTTTAG ACTGTGAAGG	1740
	CCAAATAATT TTTAAGAAAA CATTTGAAGA GTAGTGTGTT TGCAATTGTG AATAATCTTA	1800
5	CTCACAGCAA GTAAACGTAA TAAAAGCCAA CATTTAAGCC AAAAAAAAAA AAAAAA	1856
10	(2) INFORMATION FOR SEQ ID NO: 53:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1558 base pairs	
	(B) TYPE: nucleic acid	
15	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:	
20	TGGGTATCCA TTCCTGNAAT TACTTTACTT AGGATAATGG CCTCCAGCTC CGTCCAAGTT	60
	GCTGCAAAAG GTATTATTTT GTTCCTTTTT GTGGCTGAGT AGTATTCCAT GGTGTATATA	120
	TACCACATTT TCTTTATCCA CTCATTGCTT GATGGGCAGT TAGGTTGGTT CCACATCTTT	180
25	GCAATTGTGA GTTGTGCTGC TCCAGATATC ATCTTTAACT CCTTGGCCTT CTCCACATAC	240
	ATTTCCAAGT CCTGTTCATT CTACCTCCAA AATGATCTT GTATCCATTC ATCTCTCTCC	300
30	ATCTTCAATC TATTTCAATG CCCCATCATC TCTTGCATGG AGGAGTGTA TAATTGGCTA	360
	ACTGGCCTGT TCTTACATTT TAAATCAAA AGATGTGACA GGTGAAATGC CTATTTCACT	420
	GTCCATTGAT GGTTCCTGCTT ACACACCACC TGGCTGCCTG GTGTCCAGT GGCAGAGTTG	480
35	AGCAGTGTA AAAAGACTGC TTGGCCCTTT ACAGGGAAAG CAGGTCCACT GTGGCCTGTG	540
	AGGACGAGAG CTCTGGGCAG GCTCGGACAC TGGCAGACCC TGGTCCCTGG TGGCCAAGGC	600
40	AGCAGGTAT GTGTTTGGG TCACTCACAG GGCTCAGCAC CACTCCTCAT GGCTTCCTTA	660
	CTGTTTCGGC AGAGGCTGAC CCGCGGCTGA TTGAGTCCCT CTCCAGATG CTGTCCATGG	720
	GCTTCTCTGA TGAAGGGGGC TGGCTCACCA GGCTCCTGCA GACCAAGAAC TATGACATCG	780
45	GAGCGGCTCT GGACACCATC CAGTATTCAA AGCATCCCCC GCCGTTGTGA CCACTTTTGC	840
	CCACCTCTTC TCGGTGCCCC TCTTCTGTCT CATAGTTGTG TTAAGCTTGC GTAGAATTGC	900
50	AGGTCTCTGT ACGGGCCAGT TTCTCTGCCT TCTTCCAGGA TCAGGGGTA GGGTGCAAGA	960
	AGCCATTTAG GGCAGCAAAA CAAGTGACAT GAAGGGAGGG TCCCTGTGTG TGTGTGTGCT	1020
	GATGTTTCCT GGGTGCCCTG GCTCCTTGCA GCAGGGCTGG GCCTGCGAGA CCAAGGCTC	1080
55	ACTGCAGCGC GCTCCTGACC CCTCCCTGCA GGGGCTACGT TAGCAGCCCA GCACATAGCT	1140
	TGCCTAATGG CTTTCACTTT CTCTTTTGTG TTAAATGACT CATAGGTCCC TGACATTTAG	1200
60	TTGATTATTT TCTGCTACAG ACCTGGTACA CTCTGATTTT AGATAAAGTA AGCCTAGGTG	1260

TTGTCAGCAG GCAGGCTGGG GAGGCCAGTG TTGTGGGCTT CTGCTGGGA CTGAGAAGGC 1320
 TCACGAAGGG CATCCGCAAT GTTGGTTTCA CTGAGAGCTG CCTCCTGGTC TCTTCACCAC 1380
 5 TGTAAGTCTC TCATTTCCTA ACCATCAGCT GCTTTTAAAA TAAGATCTCT TTGTAGCCAT 1440
 CCTGTAAAT TTGTAAACAA TCTAATTAAA TGGCATCAGC ACTTTAACCA AAAAAAAAAA 1500
 10 AAAAAAAAAA AAANAAAAAA AAAAGGGGGC CGCTCTAGAG GTCCAAGTTA NGACGNGG 1558

15 (2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 948 base pairs
 (B) TYPE: nucleic acid
 20 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

25 TAAAAATCAT GCTCTGTACC ATCTCAGCG TAGTCATCAT CATCGCCGCG CAGACCACGA 60
 GAACTACTGG GATCCCTAAA AACGCCCTTG GTCCGGCCCC ACTCTGCGCC CCTCGATCTC 120
 CCAGGCTCTT TCTGCAGWCA TACCGCGGAC CCAATGGGCG CCCTGCACAC CCGTTTCTGG 180
 30 GGCCGTCAGA CTGGGATACA TCGTAAACTC CGCCTCCACG GAACGTCTCG CCTKGCGAGC 240
 AAGMTGGAA TCCAGTTCCT CAGGAACCCC TCCAAAACCC ACACCCCCAG GGACGCGCT 300
 35 TTCCGGGATC CCGGSCAAAC GCCGACCCCT CAGTCGCTCC AGGCCCCCTC ACCCTCAAAG 360
 TGTAAGCGCC CCAACCGAGC AACCTCGGTT TGGTCCCTAA AACCCCGCCT CCTCTATAAG 420
 CACCGCCCCA GCTCTGACAA AACCCCGCCT CCAGGTGGC AGGCTCCGCT TCTTTTCTTC 480
 40 TCCGGGGGT GATTAGTCC AGTGATTGGG TTTGTGGCTC CAGGCTCGC CCACAGACGG 540
 ACAGACCCCT CCTTTCTTC CGGCAAAAGG ACCGAGCCCT GGGGTAGTAA GGSCCCACA 600
 45 CTCTGTTTT TTGCAAGTAC ATTTTGTGCC YTCTCCACC CAGGTATCTG CCTATTTTCT 660
 TGCTAATCCC AGAACCTTC CTTTGTCTT TTTTAAGGAC ATTGGGAAG TTCCTGGTGT 720
 AGGACCCCTC TCCTGGGAT AAGAAACCTG CCTGTAAACG CTCTGTAAAT ACTCCCTTCC 780
 50 ACCCATCCCA GCCCTGGGC AGCGGGCAG AAGGGAATCC AGGCTATGGA CCTCCAAGT 840
 CCCCCTCCC CGCTCCCTC GCGGGCCCG CCTGTGCTG ATCTGTGTGT GAGTGTGTGT 900
 55 GAACTTCTGA AAGACAATAT TAAAGAGCT TAGTTGAAAA AAAAAAAA 948

60 (2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 990 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

10 GGGGAAGTGC AGTGACAGCA GGAGTAAGAG TGGGAGGCAG GACAGAGCTG GGACACAGGT 60
 ATGGAGAGGG GGTTCAGCGA GCCTAGAGAG GGCAGACTAT CAGGGTGCCG GCGGTGAGAA 120
 TCCAGGGAGA GGAGCGGAAA CAGAAGAGGG GCAGAAGACC GGGGCACTTG TGGGTTGCAG 180
 15 AGCCCCTCAG CCATGTTGGG AGCCAAGCCA CACTGGCTAC CAGGTCCCCT ACACAGTCCC 240
 GGGCTGCCCT TGGTCTGGT GCTTCTGGCC CTGGGGGCGG GGTGGGCCCA GGAGGGGTCA 300
 20 GAGCCCCGTC TGCTGGAGGG GGAGTGCCCT GTGGTCTGTG AGCCTGGCCG AGCTGCTGCA 360
 GGGGGGCGCG GGGGAGCAGC CCTGGGAGAG GCACCCCCTG GCGAGTGGC ATTTGYTGCG 420
 GTCCGAAGCC ACCACCATGA GGCAGCAGGG GAAACCGCA ATGGCACCAG TGGGGCCATC 480
 25 TACTTCGACC AGGTCTGGT GAACGAGGGC GGTGGCTTTG ACCGGGCTC TGGCTCCTTC 540
 GTAGCCCCTG TCCGGGGTGT CTACAGCTTC CGGTCCATG TGGTGAAGGT GTACAACCGC 600
 30 CAAACTGTCC AGGTGAGCCT GATGCTGAAC ACGTGGCCTG TCATCTCAGC CTTTGCCAAT 660
 GATCCTGACG TGACCCGGA GGCAGCCACC AGCTCTGTG TACTGCCCTT GGACCCCTGG 720
 GACCGAGTGT CTCTGCGCCT GCGTCGGGG NAATCTACTG GGTGGTTGGA AATACTCAAG 780
 35 TTCTCTGGC TTCTCATCT TCCCTCTCTG AAGGACCCAA GTCTTTCAAG CACAAGAATC 840
 CAGCCCCTGA CAACTTCTT CTGCCCTCTC TTGCCCCANA AACAGCANAA GCAGGANANA 900
 40 NACTCCCTCT GGCTCCTATC CCACCTCTT GCATGGGAAC CTGTGCCAAA CACCCAAGTT 960
 TAAGAAAAAA ATAAACTGT GGCATCTCCA 990

45

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1603 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

50 GGTGACCCA CGGTCCGGC CCGCGGCTC CGGAGCGGCT CTGCCTTCCC GAGCGCGGA 60
 CCGCGCCTG GGGGAGGAGG GCGAACGACG CGCGATGGC TCCGCGGCA CTCCCGGGT 120

60

	CCGCCGTCTT AGCCGCTGCT GTCTTCGTGG GAGGCGCCGT GAGTTCCCG CTGGTGGCTC	180
	CGGACAATGG GAGCAGCCGC ACATTGCACT CCAGAACAGA GACGACCCG TCGCCAGCA	240
5	ACGATACTGG GAATGGACAC CCAGAAATATA TTGCATACGC GCTTGTCCCT GTGTTCITTA	300
	TCATGGGTCT CTTTGGGTC CTCATTNGC CAMCTNGCTT NAAGAAGAAA GGCTATCGTT	360
	GTACAACAGA AGCAGAGCAA GATATCGAAG AAGAAAAAGG TTGAAAAGWT AGRATTGAAT	420
10	GACAGTGTGA ATGAAAACAG TGACACTGTT GGGCAAATCG TCCACTACAT CATGAAAAAT	480
	GAAGCGAATG CTGATGTYTT AAAGGCGATG GTAGCAGATA ACAGCCTGTA TGATCCTGAA	540
15	AGCCCCGTGA CCCCCAGCAC ACCAGGGAGC CCGCCAGTGA GTCCTGGGCT TTGTCACCAG	600
	GGGGGACGCC AGGGAAGCAC GTCGTGGCC ATCATCTGCA TACCGTGGC GGTGTWGTG	660
	AGAGGGATGT GTGTCACTGG TGTAGGCACA AGCGGTGGCA CTTTATAAAG CCCACTAACA	720
20	AGTCCAGAGA GAGCAGACCA CGGCGCCAAG GCGAGGTCAC GGTCTTTTCT GTTGGCAGAT	780
	TTAGAGTNAC AAAAGTGAG CACAAGTCAA ACCAGAAGGA ACGGAGAAGC CTGATGTCTG	840
25	TTAGTGGGGC TGAAACGTC AATGGGAGG TGCCGGCAAC ACCTGTGAAG AGAGAACGCA	900
	GTGGCACAGA GTAGCAGGTG AGCCGTGGTT TTGGTGACAT TGGGGCAGA GTGGTGACAG	960
	GTGAGGAGAA GGTACTTGA GCCTCCCAGG TGCTGTGGCA GCATAGGAAT GGTATTTGAC	1020
30	AGGGAAGTGG GAGAGCTTTC CTTGACCCAG GAAGACTGAG GGGGACTGAA CATGATTACT	1080
	TGTCTGCCTA GAGCTTCTTG TAAAGAAGTC ACAAACCTAG TGCCTCCAGG GGCTTGGCTG	1140
35	TGTGATAATG AGGATAGAGG ATTACTTGTG AGGCAATGTG GCATGGTGGG GATTGTGGCA	1200
	AACTAGAATT CACATCACCC ACCATATAGG GCTTGCATTA CCACGAGGCA GAAAGCACCT	1260
	AGTGTGCTG CATCTTCTTA CGCAAAAAG ACAAATCCA GACTTCTAAA ATGTAAAATC	1320
40	ACTGATTTTC GATATTGGCA GCTTACTTTT TTTTITTTAA CAACCATGCA GGCCAAATGA	1380
	CTGTAAATCT TGTACCAATT TTTAGGTAAA CTGTGACTTG AAAAGTCTG GAGCAAACAA	1440
45	ACCAATGCTT TTTCCTTTTA TTCTGTGGR AACCAGTTT CTTGTGTCA CAGTTYTGAA	1500
	AOCTCAATAC GAATATTCT CTCCACCA AATATTTGA GGCAATTGAA AAGCCACAGT	1560
50	GATTTATTC TTGATTGGC AATTTAATT TTGCAAGACA ATT	1603

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1052 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

5 TACAGCTCAG GATGCCTGTA ACATTGTCAT CTCGGGGCTT CTGGGTCTTG CTTAGCCTGC 60
 TTTTTCCTCG GAGGACTGAC CAGGGATGCG GCCCAGCAAC ATGTTACTAA ATCATACTCT 120
 CCTCCCTACC TTTCCCAGAC CTCTCACTCC TGCTGGTGT TCCAACCCGT TCTGTGGCCA 180
 10 GAGTATACAT TTTGGAACCT CTTGAGGCC ATCCTGCAGT TCCAGATGAA CCATAGCGTG 240
 CTTCAGCAGN AAGGCCCCGAG ACATGTATGC AGAGGAGCGG AAGAGGCAGC AGCTGGAGAG 300
 GGACCAGGCT ACAGTGACAG AGCAGCTGCT GCGAGAGGGG CTCCAAGCCA GTGGGGACGC 360
 15 CCAGCTCCGA AGGACACGCT TGCACAACT CTCGGCCAGA CGGGAAGAGC GAGTCCAAGG 420
 CTTCTGCAG GCCTTGGAAC TCAAGCGAGC TGA CTGGCTG GCCCGTCTGG GCACTGCATC 480
 20 AGCCTGAATG AGGCTGGCCA CTTGCCACTT TGCCCTGCCC TCTGCCTCCA GGGCTCCMCT 540
 MYCCTTCCTT TTCTTGCTGA AAGGCACCTC CTTTCTGAT AATGAATGGT GTTCCCTTTG 600
 CTTGGCTGGG GAGCCCCCA GGCCAGGTTT GCTGGCCATA GATACCTTTG GGCTGCCTGR 660
 25 GACAGGCTCC TGAGGAGGAT TGAGGGTGAA AGTCTCCAC GAGTACACTA AACCTAGGTC 720
 TGGTCACCAA TAGGGTTTGG AGAGCAAAGG GCCACAATC ATCAGCTGCC TGTCTCTTAG 780
 30 ATGCACTTTC TTTTTCACC AGCACATCCT TCAACACACA GAATTTGAGG GAAGAGTTCT 840
 CCCCAAAACC CTAGCTCTTT ACCCTTCCAT TTTAGCCTTC CACCCAGCTT CCACAAAAGA 900
 TTTGGCTCTA CTTGGAGTCT GCTAGTAAAT AACTAATAGG CAGGCAGTTA TTTGGGTAAG 960
 35 GAAAAAAGGG GTGGGAGAGA CAGAAAATTT GCCCACTGCT GCTCCTCCCC TTGGSTYTC 1020
 ACCTGGGATT TGCTATTGAA TCTCTACCCT NN 1052
 40

(2) INFORMATION FOR SEQ ID NO: 58:

45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 814 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

ACNCGNTGGC GGCCGCTCTA GAAGTAGGGG ANCCCCGGG CTGCAGGAAT TGGCAGCAG 60
 55 CATAGACTTT TAAACTGGTA CGGTCTTAG AGATGGTCCT TGGCCTTCG TTGTTGTTGT 120
 RGTTTTTTTC TTTTCTCTT TCTCTCTC CTCTCTCTC TCTCTCTT CTCTCTCTT 180
 TTTTTTTC GAGTCTGCT CTGTACCAA GACTGGAGTG AAGTGATGTG ATCTCGGCTT 240
 60

ACTGCAACCT GGGAGGCAGA GGTTCAGTG AGTCGAGATG GTGCCATTGC TCTCGTTTGG 300
 GCAACAAGAG TGAAACTCTT GTCTCAAAAA AAAAAAAAAA ATGAGGTTTA AGACAGTTTT 360
 5 GTCAATTACTG GTGGGATCTG GTCACACAAG ATAGCATTA ACGTGACATG GCACATAAAA 420
 TTGGTTAAAA AATTTTGT TTAAATTACG TAATGTAAAA GCCCAACAAA CACTTTATGC 480
 AAGATTGAA TGTATCTTCA AATTCAGATT TAATAACAT GTAAAGATCC TCTGTATATA 540
 10 AAAGTTGTAT TTAATCCCTT GTGCCCAAG AATGCTATAA AAGATCCCAA GAATGTTATC 600
 TATGAAAAGA TAGCAATAGG GAATGGTGAA CAAATAATTT AATTTGCCAA TTCTAAAAAA 660
 15 CATGGACTTA AACCCCATGA AACTTGGTT CCATAGTTTT AACTGTTTTA TGGTTCCAAT 720
 ACAAACCCAG AGTGGTTTAC ATTCCACAAT NACCAAATTT GCATCCAATN TTGGGGTAAT 780
 20 TTTNGGTATT TGCCATGGGA TACTATTTCAT TTTT 814

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1215 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

AGAGGAAGTC TTTTGCCAAG CCTGTCTCT GGAATAACGC CATCCAGGCT GGGAGGGGAA 60
 35 GAGTGTCTG CTACACTGCT CCCCTCTCT CCTCATCTTC CTTCTCAGCC TTGGTTCTCTG 120
 ATGGGAACAG AATGGAGGGC CTGAGAACAT ACTTTCTAAA TGCCTTTGAC CCAGGAACCG 180
 40 ATTATCTATA TTTGTTCCCA TTTTCTCTCA CCGTGACATT CCAGCATTTG CTGACTGTGA 240
 GGTGGGCTT TGAGAGCTC CAGGTTCTC AAAACAGGCC TGAGCGATGG GCATCACACC 300
 45 CTCTGCCTAC CCACRTGCT GCTTACCTGC CAGATAACCA AGTGNAGATG TCTGCCAGTG 360
 GCTAGTTTTC ACATTCTTAC TAGTGTTTGG YTCACCTTTG GGCAAAGGCC CCCTCTAGGC 420
 CTTGCCCCAC CTCCATCAAA CGCAGACACT GTAGTCAGAC CTCAGYAATA TAGGAGGCAA 480
 50 TAATCTTTTA ACAGTGTMTT GCAAACAAAC AAAAAGAGAA AAATCCCAGC CAGGGGAAT 540
 CGCCACCTGC CCACGCTAGT TOCATCCAG CTCAAGACCC GCCCTTAGAC CAGGCAGGCA 600
 AAGGCCCCCA TCACACTCGG CCACTAGTGG GGTCTGAGG CCAAGAAAGA AACCAGACCC 660
 55 TGTATGACAA GTTGGGKTCT TTCCAGAACA CGACAGAAAC AGGGGGGGCC CTTGTTAAT 720
 GCCACTCCAT ACTCCAGAAG CATTATCTCT TATTTGGGAC AGCCAAGGC AGATTACAG 780
 60 GTTATTGTAG GAATAAGAC TAGTTTACAA AGGARAAGA GSCCTGGAC TTCCCMAGGA 840

5 AAGGTCAGGT TAGGGCTCCT GTACCCATTC TGTTCACCA CTGTTTGATC TCTCTGGCCT 900
 CCCACCAGGA ATGCCGTTTC CTTTATATGG ATCTGTGTTGG AACAGAGAG AATCAACAGA 960
 TCAATGACAT AGGATCOGAA GTGCAATGAT AGTCACCTCT AGTTTGGCAT TTCACAACT 1020
 CTGNACAGCA AGGTATTTGGT AGGTTACTCA ATTTCAAAG GCCCCATGG CCAAATATGT 1080
 10 TTAGGAACCG CTGTTTGNAT TTCTTTTTTT GGAGACGCAT TGTATATAAT ATATGTCAA 1140
 GGCTTTCGGA ATTCCTGCAG GAAAGAAATC AGCTTTGTTA AATCCNAAAA AAAAAAAAAA 1200
 AAAAAAATAG ACTCG 1215
 15

(2) INFORMATION FOR SEQ ID NO: 60:

20

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 478 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

30 ATTTCTTATG ACATGGGGGT TTGAATTGGT TGGCAAATGT TTAATTTTAA TATCCATAAT 60
 CAGTGAGGTC CTGCTGGCTG TAATCATTA TGTGAAATC TAAGGAGCTT AGTTCATGGC 120
 TCTAGAATT CACAGAAAR TGYGMTATGA TACGAGCATT AAGTTTATTT CTTCTGATCT 180
 35 TTGATGCAGC TTGTTTCAGT TTAATCTGTTT TTGATTTTAT TGGTCATCTA CTTCCCATGC 240
 CAAAAGGGAC TGGTCTACAT AGCTGCGCTA AACACCTGAT CAAATCACTA AAAGAAAATG 300
 TGTTACCTCT AATGAATTAT CCTGATTGTA AGTTAAAAAT CAATATTTCC CCGTAGTGAG 360
 40 GTTTGCTTTT TAAAAAGAA KCTTAAAAAA AAAAAAAAAA AAACGAGTTN AAGAAAAGGA 420
 AGCAAGCTCA GGTAAGGTGC ACACATTGGG CTAAGGAAGC TAGAGCCTGT GGAGANGC 478
 45

(2) INFORMATION FOR SEQ ID NO: 61:

50

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 618 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

TATGACCTTG ATAACCCCAA GTTNGAAATT AACCTTCANT AAAGGGAACA AAAGCTGGAG 60
 60 TTCGCGCGCT TGCACTTGA CACTAGTGGA TCCCAAAGAA TTCGGCACGA GTCATAATGA 120

320

5 GCTACTAGGT AAGCCTTCTG GGACTTTCAG ATATTTTGGG GAAGATTGAT TTTTGTTCCTT 180
 ACATGCTGTG GACCCCTGGC CATCAAATGG TATGGGGAAG CTCATCCGTC TGTCTGTGAT 240
 GGTCACTGCA GTCAGGCGTC TTTTITAGTAT TTA CTGGGTG CTCAGTACTG TGCCAGATGC 300
 TGTCCGGAGC CGTGGTGGTA TGGAGGAGGA GTGCTCCAGA GGACTCTGCT GTGTGGCAGG 360
 10 CCAGCATAAA CAAGCCAAGG GGAAAAGGCA GGCATGGAAT AAAGGGGGAG AATACCAGTG 420
 TGTGACTTAC TGCTGACTGT GTGGATTAGC CTATCAGCAG TAATCAAGCA GGGCGGAGGG 480
 CATTTATCTTT GAGCCAGAAG AGTGAGCACT GGSCCGAGGG TGGAGCATCA AGAGGGGGTG 540
 15 TAGGACCNCA AGGCTTCTTN CNGGGGAGAC AACGTCAATA AGCNGTCAGT AGTCACCGAC 600
 AGTTTGTGGA AGCAAGGG 618

20

(2) INFORMATION FOR SEQ ID NO: 62:

25

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 751 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

TCGACCCACG CGTCCGAGGA GCTGGACTTC TGAGACAGCC ATTCTCCTTG CATAGCACTG 60
 35 TCTGCTGCTA CAGCTCATAG AAGTCAACAA TTTTCTTCAA CACTGGTAGG CAGCCTCTAA 120
 ATGGCCCTGA TCACCTTCAC CTCTGCCAT TCACACCNNT GTAAAATTC ACCCCTGGAC 180
 CTAGTGACTC ACTTCTAACA ANGAGAATAC AGCAAAAGTA ACATCGCTTC TGAGGTGAGG 240
 40 CTACAAGGAG ACTACGATGC CTGCCTTGGT CACCCCTTCTC CTGCTCTTTC CATTGCTCCC 300
 TCTGATGGAA GCCAGTTGCC ATGTGATGAG GTGCCCTATG GAGAGGCCCA CGTGACAAGG 360
 45 TATTGTAAAA AGCCTCTGAC CAATAGCCAT CTAGAAACGG AGGCCAGTC CAGCAGCCTC 420
 TGAGATGAAT CCTGCCAACC TGAGCTTGGA GACAGATTCT CTCCTATCC TGCCTTGGA 480
 TGATCACAGC CACCACCAAC ACCTTCACTG CCTGGTGAGA GGCCAAGCCA GTGAACCCAA 540
 50 GGTAAACTGG ACAGAATCCT GACCCACAGA AACTGAGATA ATGTTTGTTA TTTTAAGCTG 600
 CTCAGTTTGT TACAGAGCAA TAGATACTA ACTCAAACAC CATAAAATTC TAATATTTTA 660
 55 TTCTATCACA CAAACCAGGT AATACCAAGT AAATGCCATT ACTATACACA TATTTTGTGA 720
 ACACAATTAC ATGTGATTTT TTAAGAAGGC T 751

60

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 780 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

CNNGCAGTCA CAGTCCCGGA TTCCCGGGTC GACCCACGCG TCCGGGTGG CAACTCCTGA 60
 GGCCTGCATG GGTGACTTCA CATTTTCCTA CCTCTCCTTC TAATCTCTTC TAGAGCACCT 120
 GCTATCCCCA ACTTCTAGAC CTGCTCCAAA CTAGTGACTA GGATAGAATT TGATCCCCTA 180
 ACTCACTGTC TCCGGTGCTC ATTGCTGCTA ACAGCATTGC CTGTGCTCTC CTCTCAGGGG 240
 CAGCATGCTA ACGGGGCGAC GTCCTAATCC AACTGGGAGA AGCCTCAGTG GTGGAATTCC 300
 AGGCACTGTG ACTGTCAAGC TGGCAAGGGC CAGGATTGGG GGAATGGAGC TGGGGCTTAG 360
 CTGGGAGGTG GTCTGAAGCA GACAGGGPAT GGGAGAGGAG GATGGGAAGT AGACAGTGGC 420
 TGGTATGGCT CTGAGGCTCC CTGGGGCCTG CTCAAGCTCC TCCTGCTCCT TGCTGTTTTC 480
 TGATGTTTTG GGGGCTGGG ASTCCCTTG TCCTCATCTG AGACTGAAAT GTGGGGATCC 540
 AGGATGGCCT TCCTTCTCTT TACCCCTTCCT CCCTCAGCCT GCAACCTCTA TCCTGGAACC 600
 TGTCCCTCCT TCTCCGCCAA CTATGCACT GTTGTCTGCT CCTCTGCAA GGCACGCCAG 660
 CTGGGAGCA GCAGAGAAAT AACAGCAAT TCTGATGCCA AAAAAAAAAA AAAAAAACC 720
 GCGGCGGAAA GCTTATTNCC CTTTAAGTAA GGGGTTAATT TTTAGCTTGG GCACTNGGCC 780

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 588 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

TTCCGAAATTA ATGACTCAC TATAGGAATW GCGTGGCCA TGACCCGCGG TAACCAGCGT 60
 GAGCTGGCCC GCCAGAAGAA TATGAAAAAG CAGAGCGACT CGGTAAAGG AAAGCGCCGA 120
 GATGADGGGC TTTCTGCTGC CCCCCGAAG CAGAGGACT CGGAGATCAT GCAGCAGAAG 180
 CAGAAAAGG CAAACGAGAA GAAGAGGAA CCAAGTAGC TTTGTGGCTT CGTGCCAAC 240
 CCTCTGCCC TTGCGCTGTG TGCTGGAGC CAGTCCACC ACGCTCGCGT TTCTCTCTGT 300

322

AGTGCTCACA GGTCCCAGCA CCGATGGCAT TCCCTTTGCC CTGAGTCTGC AGCGGGTCCC 360
 TTTTGTGCTT CCTTCCCCTC AGGTAGCCTC TCTCCCCTG GGCCACTCCC GGGGGTGAGG 420
 5 GGGTTACCCC TTCCAGTGT TTTTATTCC TGTGGGGCTC ACCCCAAAGT ATTAAAAGTA 480
 GCTTTGTAAT TCCAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 540
 AAAAAAAAAA AAAAAAAAAA AAAANNCGGG GGGGGGCCCC CCCCCCCC 588

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 774 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

TTTAAAGATG AAGAAATGAC AAGGGAGGGA GATGAGATGG AAAGGTGTTT GGAAGAGATA 60
 25 AGGGGTCTRA GAAAGAAATT TAGGGCTCTG CATCTAACC ATAGGCATTG TCGGGACCGT 120
 CCTTATCCCA TTAAATTAAT TTCTCTGACA ATTCAATTAT TTCTGTTAT TAATGTTGCC 180
 30 ACTGCTTTCT GTTGTCTGC ACTTCTTGA TAAATATTG CTATCGTTTT ACTCCAGTCA 240
 TTCGATGTTG CTGAGATTTA CATATGACTC TTGTCAACAT CTCATCTTTT GACCCAATCT 300
 TATTCATTTA ATAAGAGGTC TCATTCATTT GCATGGAAAA ATGCTCATTG TATATTGCAA 360
 35 AGTGAAAATA ACGAGTTGCA AACAGTGTA TACATATATG TGTGTATATA TGTACACTTT 420
 ATTGTACAT TTCTATGTGA CATAATGCAA AGGAAAGTGT CTGATTTTAT TATACACCAA 480
 40 AGGTTAACAG TGAATCTCTG TGTGATCTCT TTTTITTTCT TTTTGCCTAT CTGCATCTTC 540
 TCACTTGCCA AAAAATGAAT ATATGTTTAT GTGTGTATAT TACTTGTGTC ACAAAAAACC 600
 CTAAAGTAGA CAGTAAAGA ACTTGTCAAT GGCCTTTGGA AGGCAATGAA ACACTTAATA 660
 45 AACTCTCAAT AACAGAAGCG TAAAAATGAA ATGTAAACCT CCAATTACCT CTGGATCTCT 720
 TAGCCAGAGT AATAAATGG TAATTATTAC AGATAAAAAA AAAAAAAAAA AANA 774

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1866 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

	ACCCACGCGT CCGTCCTCT TCTTCAGCAC ATGCCAAAGC TGTTCCTCAC GGCTGTGAG	60
5	ACAAGAGCAT CTTGGATGTA GGACAATGGA AGAGTTAGAT GCCTTATTGG AGGAACTGGA	120
	ACGCTCCACC CTTCAGGACA GTGATGAATA TTCCAACCCA GCTCCTCTTC CCCTGGATCA	180
	GCATTCCAGA AAGGAGACTA ACCTTGATGA GACTTCGGAG ATCCTTTCTA TTCAGGATAA	240
10	CACAAGTCCC TTGCCGGCGC ATTCGTGTAT ACTACCAATA TCCAGGAGCT CAATGTCTAC	300
	AGTGAAGCCC AAGAGCCAAA GGAATCACCA CCACCTTCTA AAACGTCAGC AGCTGCTCAG	360
15	TTGGATGAGC TCATGGCTCA CCGACTGAG ATGCAGGCCA AGGTTGCACT GAGAGCAGAT	420
	GCTGGCAAGA AGCACTTACC AGACAAGCAG GATCACAAGG CCTCCCTGGA CTCAATGCTT	480
	GGGGTCTSG AGCAGGAATT GCAGGACCTT GGCATTGCCA CAGTGCCCAA GGGCCATTGT	540
20	GCATCCTGCC AGAAACCGAT TGCTGGGAAG GTGATCCATG CTCTAGGGCA ATCATGGCAT	600
	CCTGAGCATT TTGTCTGTAC TCATTGCAAA GAAGAGATTG GCTCCAGTCC CTTCTTTGAG	660
25	CGGAGTGGCT TGGNCTACTG CCCCACGAC TACCACCAAC TTTTCTCTCC ACGCTGTGCT	720
	TACTGCGCTG CTCCCATCCT GGATAAAGTG CTGACAGCAA TGAACCAGAC CTGGCACCCA	780
	GAGCACTTCT TCTGCTCTCA CTGCGGAGAG GTGTTTGGTG CAGAAGGCTT TCATGAGAAG	840
30	GACAAGAAGC CATATTGCCG AAAGGATTTC TTAGCCATGT TCTCACCCAA GTGTGGTGGC	900
	TGCAATCGCC CAGTGTGGGA AAATACCTT TCAGCCATGG AACTGTCTG GCACCCAGAG	960
35	TGCTTTGTTT GTGGGGACTG CTTCAACAGT TTTTCTACTG GCTCCTTCTT TGAATGGAT	1020
	GGACGTCCAT TCTGTGAGCT CCATTACCAT CACCGCGGG GAACGCTCTG CCATGGGTGT	1080
	GGGCAGCCCA TCACTGGCCG TTGTATCAGT GCCATGGGGT ACAAGTTCCA TCCTGAGCAC	1140
40	TTTGTGTGTG CTTTCTGCTT GACACAGTTG TCGAAGGGCA TTTTCAGGGA GCAGAATGAC	1200
	AAGACCTATT GTCAACCTTG CTTCAATAAG CTCTTCCAC TGTAAATGCCA ACTGATCCAT	1260
45	AGCCTCTTCA GATTCTTAT AAAATTTAAA CCAAGAGAGG AGAGGAAAGG GTAAATTTTC	1320
	TGTTACTGAC CTTCTGCTTA ATAGTCTTAT AGAAAAAGGA AAGGTGATGA GCAATAAAG	1380
	GAACCTCTAG ACTTTACATG ACTAGGCTGA TAATCTTATT TTTTAGGCTT CTATACAGTT	1440
50	AATTCTATAA ATTCTCTTTC TCCCTCTCTT CTCCAATCAA GCACTTGGAG TTAGATCTAG	1500
	GTCTTCTAT CTGTCCTC TACAGATGTA TTTTCCACTT GCATAATTCA TGCCAACACT	1560
55	GGTTTCTTA GGTTTCTCCA TTTTCACCTC TAGTGATGGC CTTACTATA TCTTCTCTAA	1620
	TTTGGTCTG ATACTTGTTT CTTTTCACGT TTTCCCATTT CCTGTGGCT CACTGTCTTA	1680
60	CAATCACTGC TGTGGAATCA TGATACCACT TTTAGCTCTT TGCATCTTCC TTCAGTGTAT	1740

TTTTGTTTTT CAAGAGGAAG TAGATTTTAA CTGGACAACT TTGAGTACTG ACATCATTGA 1800
TAAATAAACT GGCTTGTTGGT TTCAATAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1860
5 AAAAAA 1866

10 (2) INFORMATION FOR SEQ ID NO: 67:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1152 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

20 CTCAAGGATG TAAAGGCTCT GCAGATTTTCG GGAGGCCTGT CTCCCAGCAC CTGATGGGAC 60
ACTTTTGGCC CCACTGTAAA TTCTGGGTGT ATCCTCCACT GTATGCTGTC ACCCCAAGGG 120
25 CAAGCACTGC ATCTGCTTAG TGAAGGATTT ATTGTTGGGA AGATACATTT TCCCCTTKAG 180
CAGAGAGTGG CGTATCCTGG CAGTCTTCGG TGAGCCAGTT GTACCAGGAT TATGAAATGC 240
AGATGTTTAC TGTTTCATTG TTGCTGTCAT TGCTACTGAG GAGTACTGAC CAGAATCATC 300
30 TGCAACTYTT AGTTGGCAGA GAGGACCACT ATGGCGGGTA GCTCTTTTCT TTCTTGCCAT 360
TGTGGGGATG ATTCCAGGCC AAAGATGATG GARAAGTATG GAAATCATCT GAAAGGTTGA 420
AGCTTGGCAC GTGAAGCCAT TCATGACTTT GTAAGGCAGT TTTGCTGAAG GCCAGTTCTG 480
35 CCTGGGAGG GACGGAGGTG AATCCTCCTG AGTACCTGTG GTTTTCTTAC TTCTGCTGA 540
ATTTACCTAA GTGCCTGTTG TTTGCTTGCT GTGGAGGCTT TCTGGTATTT CATTTTCAGGT 600
40 GCAGATGCCT TCACTTTCCC ACCRAAAAAA CCCCMACCAA ACCTAAGACC TTAAGTCAAC 660
TAAGTYTNCC AAGTACTTTT TAACCCAATG GGATGAACAG CCTGTGGTCT GCTCAGATCA 720
CCTGAGTGC GTGTGAGAAG GCMINGGCTT TGCCAGGAAA TCCAGGAAGG CAGGGCCGGG 780
45 CTGTGTTGGA AGCTGGCTTA GCTGGTGGGG CAGCCTTATT TCAATTAAAA GGGCATTGAC 840
TGGGAGCAGC AGTCTGGAG TTGTTGTCAT TTCCTATTGC CCTCAAAATG AGAAACCAGG 900
50 AAAATAGCAG ATTGGAGCCT TCGAGAAGGC AGTAAATGGC TGTTTTATT GACAAAAGGA 960
AAACATTTTA CTGCCATCTC ACTGATGGCA TCTCACTGAC TTAAATGAA GGCANGTTGT 1020
AGTAAAAAAA AAAGTCTACA TTTTCCACC GCCACGTTCT TATATCCTGT TTGTCAGCCA 1080
55 CTGCTCANAA GGGCATGTTG TCTTGCGGAN TANAGGCGCT CTCCTCCCT CGTTTCCCT 1140
ATAGGTGGG TG 1152

60

(2) INFORMATION FOR SEQ ID NO: 68:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2483 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
10 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

AGCAGGCGGT GCGCTGGGGG CGGAGCAGC GCGKAGCCCG GCTCGGCCAC ACCGATCGCC 60
15 CGCCGCCATG GGCTCCTCGC AAAGCGTCGA GATCCCGGGC GGGGGCACCG AGGGCTACCA 120
CGTCTGCGG GTACAAGAAA ATTCCCAGG ACACAGAGCT GGTTTGGAGC CTTCTTTGA 180
TTTTATGTT TCTATTATG GTTCAAGATT AAATAAGAC AATGACACTC TTAAGGATCT 240
20 GCTGAAASCA AACGTTGAAA AGCCTGTAAA GATGCTTATC TATAGCAGCA AAACATTGGA 300
ACTGCGAGAG ACCTCAGTCA CACCAAGTAA CCTGTGGGGC GGCCAGGGCT TATGGGAGT 360
25 GAGCATTCGT TTCTGCAGCT TTGATGGGGC AAATGAAAAT GTTTGGCATG TGCTGGAGGT 420
GGAATCAAAT TCTCTGCAG CACTGGCAGG TCTTAGACCA CACAGTGATT ATATAATTGG 480
AGCAGATACA GTCATGAATG AGTCTGAAGA TCTATTGAGC CTTATCGAAA CACATGAAGC 540
30 AAAACCATTG AAACGTGATG TGTACAACAC AGACACTGAT AACTGTGAG AAGTGATTAT 600
TACACCAAAT TCTGCATGGG GTGGAGAAGG CAGCCTAGGA TGTGGCATTG GATATGGTTA 660
35 TTTGCATCGA ATACCTACAC GCCCATTTGA GGAAGGAAAG AAAATTCTC TTCCAGGACA 720
AATGGCTGGT ACACCTATTA CACCTCTTAA AGATGGGTTT ACAGAGGTCC AGCTGTCTCTC 780
AGTTAATCCC CGTCTTTGT CACCACCAGG AACTACAGGA ATTGAACAGA GTCTGACTGG 840
40 ACTTCTATT AGCTCACTC CACCAGCTGT CAGTAGTGT CTGAGTACAG GTGTACCAAC 900
AGTACCGTTA TTGCCACCAC AAGTAAACCA GTCCCTCACT TCTGTGOCAC CAATGAATCC 960
45 AGCTACTACA TTACCAGGTC TGATGCCTTT ACCAGCAGGA CTGCCCAACC TCCCCAACCT 1020
CAACCTCAAC CTCCAGCAC CACACATCAT GCCAGGGTT GGCTTACCAG AACTTGTAAG 1080
CCCAGGTCTG CCACCTCTTC CTTCCATGCC TCCCAGAAAC TTACCTGGCA TTGCACCTCT 1140
50 CCCCCTGCCA TCCGAGTTCC TCCGTCATT CCCCCTGGTT CCAGAGAGCT CTTCTGCAGC 1200
AAGCTCAGGA GAGCTGCTGT CTTCCCTCCC GCCACCAGC AACGCACCT CTGACCCTGC 1260
55 CACAACACT GCAAAGGCAG ACGCTGCCTC CTCACCTACT GTGGATGTGA CGCCCCCAC 1320
TGCCAAGGCC CCCACCACCG TTGAGGACAG AGTCGGCGAC TCCACCCAG TCAGCGAGAA 1380
GCCTGTTTCT GCGGCTGTGG ATGCCAATGC TTCTGAGTCA CCTTAACCTT GAACCATCT 1440
60

	TTGGAATTGG CGTGGTATAT TTAACCACGG GAGCGTGTCT GGAAACGCAA ACTATCATT	1500
	ATTCATACT AGTTTGTACC GTATCTGTAG GCATCCTGTA AATAATTCCA AGGGGAAAAC	1560
5	TAAACGAGGA CGTGGGTGTG ATCCTGCCAG GTTGAGTGGG GCTCACACGC TAGGGTGAGA	1620
	TGTCAGAAAAG CGCTTGTATT TTAACAACCC AAAAAGAATT GTAAGGGTGG CTGCTGCCA	1680
	GGCTTGCACT GCGTTCCTG GGGGTGTGCA TCTTCGGGAA AGGTGGTGGC GGGGCGTCCA	1740
10	CTAGGTTTCC TGTCCCTGC TGCTCCTTC GTAGAAAAT GAAATATTCT ATGCCTAATA	1800
	CTCACACGCA ACATTTCCTG TACTTTGTAA GTCGTTTGGC AGAATGCAGA CCACCTCACT	1860
15	AAACTGTAAA CGGTAAAGAG ATTTTACTT TTGGTCTCCG TGAGTCGCAT CTCTACTAAG	1920
	GTTTACACAG GAATTCACCC TGAAGACTTG TGTAAAGTT CTACAGCGCG CACTGTTAAC	1980
	TGAACGTCTT TTTCTTCAGC CTATACGCGG ATCCTTGTTC TGAGCTCTCA GAATCACTCA	2040
20	GACAACATTT TGTAACGTCT GCTGTGCTT TCTACATACA CCTTATAAAG TGACATTTCA	2100
	AAAGAAATAA GGTGCCACAG TTTTAAACCA GAAGGTGGCA CTCGTGGCT CCTTGTAGTA	2160
25	TTATAGCTAT ACTGGGAAAG CATAGATACA GCAATAAAGT ACAGTAATTT TACTTTTTTT	2220
	CTTGTTTAC ATCTAAATTA CAACCTTAA TTGCCAGTG TGCCTTACT ACTCTCCAGT	2280
	ATGTCTTATT ACTCTCCAGT ATGTCACGCA TCTTTAAGT TTCAGTCTT ATGTTTGTCT	2340
30	TCTCCCATTT TTAAGAGATG GTAAGTTAAC TGAATTGAT TTACTGAATG AAATTAAATG	2400
	CAGATATCCC TGTMTTGGAA ATAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	2460
35	AAAAAAAAA AAAAAAAAAA AAA	2483

40 (2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 536 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

50	GAGAAATGGA GCTTTGTAG ATAAAAATTT TTTCAACGCA AACAGTCATT TTCCAGTGAA	60
	AGGAGAGCGT ATCCGCGTA GGATGGACTT AGATCGTGTA AAAGCTGAGG CCACCGAGGA	120
	TATAACCTCC GGGTCCCTTT GCCTCCTTTT CCTTAGACTC CCTCCAACT CGTGATCTT	180
55	TCCTTCAGCA GTACTGGGCT CCACCGCAAC CTAGTCCTTT GTCTTTACCC TATTACCTTT	240
	CATAACATCC TAGTTGAAAA GTARTTATTC AACCGCGTTT GAAAATGAGA ACAGGTTTAC	300
60	AGARGCTAGG TTACTTGCGA AGTTCGTCA ATTAGTAACC AGTAACGCCA GGAAGTCCAG	360

327

TTTCTTGCTT CGAATTCTC ATGGTAGCTT TCACCARGCT CCCCGTCMAA TGCTAACGTC 420
AACTACTGAA CTAGATTAGC AAAAAGGTCT TTAAACAGAA TTCCTGGTTT TCAGAGAGAG 480
TTTCTTTTCAT GAAGCGCCCC ATTCTACAG AGGAAAATAA ACTCCAAGCA GCCAGT 536

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 865 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

CCACGGGTCC GGCTTTCTT GCCCAGAGGC GCCGGTGGGA CTCACGGGCG GGGCATGATG 60
GGTAACAGGA CCGTGGGGT CCCAGGAAG TCCTAGAGGG GGTCCGGTT TGGGTGGACA 120
AGCTTTCTCT GTCTCTCCC GACAGAGCTG ACGTGTCTG GGTCCACCG GGAGCGGCA 180
TTTCCACCGG ACGGAGGGT TCGGGGTGTC CGGGGCTGGG GAATACGTAG GGGTTGCCGC 240
GCGGTGTGGG GAGTTGGGGC GTGTGGCTGC AGTCCCGGA GTTCTTGAG GGGTCCGCC 300
CACCGAGCTT CCGGACCGC TGATCTGCCC GTAGCTTGCC GGANGARGG CGGAGCTGAC 360
TCTCCGTCCC TTCTCCATC COCTCCAGTG GTGGGTACGG GCACCTCGCT GCGCTCTCC 420
TCCCTCTGT CCTGTCTGT CTTTGTGGG ATGCAGATGT ACAGCCGTCA GCTGGCTTCC 480
ACCGAGTGGC TCACCATCCA GGGGGGCTG CTTGGTTCGG GTCTCTTCGT GTTCTCGCTC 540
ACTGCCCTCA ATAATCTGGA GAATCTGTG TTTGGCAAAG GATTCCAAGC AAAGATCTTC 600
CCTGAGATTC TCCTGTGCCT CTTGTGGCT CTCCTTGCAT CTGGCTCAT CCACCGAGTC 660
TGTGTACCA CCTGCTTCAT CTCTCCATG GTTGGTCTGT ACTACATCAA CAAGATCTCC 720
TCCACCCTGT ACCAGGCAGC AGCTCCAGTC CTCACACCAG CCAAGGTCAC AGGCAAGAGC 780
AAGAAGAGAA ACTGACCTG AATGTTCAAT AAAGTTGATT CTTGTAAAA AAAAAAAAAA 840
AAAAAAAAA AAAAAAAAAA AAAAA 865

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 932 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

5 TCATCATATA CAAAGTTTTT CGTCACACTG CAGGGTTGAA ACCAGAAAGTT AGTTGCTTTG 60
 AGAACATAAG GTCTTGTCGA AGAGGAGCCC TCGCTCTTCT GTTCCTTCTC GGCACCACCT 120
 GGATCTTTGG GGTTCGCCAT GTTGTCACG CATCAGTGGT TACAGCTTAC CTCTTCACAG 180
 10 TCAGCAATGC TTTCCAGGGG ATGTTCAATT TTTTATTCCT GTGTGTTTA TCTAGAAAGA 240
 TTCAAGAAGA ATATTACAGA TTGTTCAAAA ATGTCCTCTG TTGTTTGGGA TGTTTAAGGT 300
 AAACATAGAG AATGGTGGAT AATTACAACCT GCACAAAAAT AAAAATTCCA AGCTGTGGAT 360
 15 GACCAATGTA TAAAAATGAC TCATCAAATT ATCCAATTAT TAACTACTAG ACAAAAAGTA 420
 TTTTAAATCA GTTTTCTGT TTATGCTATA GGAAGTGTAG ATAATAAGGT AAAATTATGT 480
 20 ATCATATAGA TATACTATGT TTTTCTATGT GAAATAGTTC TGTCAAAAAT AGTATTGCAG 540
 ATATTTGGAA AGTAATTGGT TTCTCAGGAG TGATATCACT GCACCCAAGG AAAGATTTTC 600
 TTTCTAACAC GAGAAGTATA TGAATGTCCT GAAGGAAACC ACTGGCTTGA TATTTCTGTG 660
 25 ACTCGTGTG CCTTGAAAC TAGTCCOCTA CCACCTCGGT AATGAGCTCC ATTACAGAAA 720
 GTGGAACATA AGAGAATGAA GGGGCAGAA ATCAAACAGT GAAAAGGGAA TGATAAGATG 780
 30 TATTTTGAAT GAAGTGTTTT TTCTGTAGAC TAGCTGAGAA ATTGTTGACA TAAATAAAG 840
 AATTGAAGAA ACACATTTTA CCATTTAAAA AAAAAAAAAA ACTNGAGGGG GGCCCGGTAC 900
 35 CCAAATCGCC GCATAGTGAT CGTAAACAAT CT 932

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 996 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

50 CGCCTGGCAC CATGAGGACG CCTGGGCTC TGCTGTGCT GCTGCTGCTC CTGGCGGGAG 60
 CCCCCGCCG GGGGCCACT CCCCCGACCT GCTACTCCCG CATGCGGGCC CTGAGCCAGG 120
 AGATCACCCG CGACTTCAAC CTCTGCAGG TCTCGGAGCC CTCGGAGCCA TGTGTGAGAT 180
 55 ACCTGCCAG GCTGTACCTG GACATACACA ATTACTGTGT GCTGGACAAG CTGCGGGACT 240
 TTGTGGCTC GCCCCGTGT TGGAAAGTGG CCCAGGTAGA TTCCTTGAAG GACAAAGCAC 300
 60 GGAAGCTGTA CACCATCATG AACTCGTTCT GCAGGAGAGA TTTGGTATTC CTGTTGGATG 360

329

ACTGCAATGC CTTGGAATAC CCAATCCCAG TGA CTACGGT CCTGCCAGAT CGTCAGCGCT 420
AAGGGAAC TG AGACCAGAGA AAGAACCCAA GAGAACTAAA GTTATGTCAG CTACCCAGAC 480
5 TTAATGGGCC AGAGCCATGA COCTCACAGG TCTTGTTGTTA GTTGTATCTG AAAC TGTTAT 540
GTATCTCTCT ACCTTCTGGA AACAGGGCT GGTATTCCTA CCCNGGAACC TCCTTTGAGC 600
ATAGAGTTAG CAACCATGCT TCTCATTCCC TTGACTCATG TCTTGCCAGG ATGGTTAGAT 660
10 ACACAGCATG TTGATTTGGT CACCTAAAAA GAAGAAAAGG ACTAACAAGC TTCACTTTTA 720
TGAACAATA TTTTGAGAAC ATGCACAATA GTATGTTTTT ATTACTGGTT TAATGGAGTA 780
15 ATGGTACTTT TATCTTTTCT TGATAGAAAC CTGCTTACAT TTAACCAAGC TTCTATTATG 840
CCTTTTCTA ACACAGACTT TCTTCACTGT CTTTCATTTA AAAAGAAATT AATGCTCTTA 900
AGATATATAT TTTAYGTAGT GCTGACAGGA CCCACTCTTT CATTGAAAGG TGATGAAAAT 960
20 CAAATAAAGA ATCTCTTAC ATGARAAAAA AAAAAA 996

25

(2) INFORMATION FOR SEQ ID NO: 73:

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 785 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

40

GGCAGAGGG GCTTTGCGTA CACAATAGCT GCTAGGAGTA CCCAAAGCCT GARTACARCC 60
TGCTGGTGTC ATGGCCACGT GTGAGCAGGC CAGCGTCAMA CGGCTCGCTG TGACCCGTCC 120
CGRAGACTGA AATGGCCTG GGTCTTCTCC TKGTCCTGTG ATWAAAGTCC TCTCTTGAAA 180
GTGGAGAGCA AAGGCACACA GAGGTGCGCG CTCACAAGAA TTCCTCCCGG TGACTGGGTA 240
ATCAATGTTA CTGCTGTTTC CTTTGCAGGA AAGACCACAG CAAGATTCTT TCAITCGTCT 300
45 CCTCCTAGCC TGGGGGACCA GGCTCGAACT GACCCTGGAC ATCAAAGGAG GGATTATGTG 360
GCTGTCAAAG CCATCGGCCC ACAGCCCTGT TCACRTCTTG GTGCTTCTCT TCCCAGAGG 420
50 CTGGTCCCAG CCAGGCACAC ACAAAGGCA GATTCTCGTA AACSCAGCCT CCTCCCTGG 480
AGGCTGCCTC CTGCCCTGGA TCTGGAGTGG AGCTGCTCTG AGATTTTGAG TTCTTCTGCA 540
GAGATGATTA AATATATCCA AGAGACATG GAAACCTGC TGAACATTTT ACATTGGTCT 600
55 GCTCAGCACA TGGCTGGATG CGGATATTTT TATAATTCCA GAAAGTCACA CAGCTCCTCT 660
GTATGAGACC AGTGGGCGCC ATTTAAAAGA ACAGGATGAG AATCTAAGAT ATATTATTAA 720
60 TAAATGTAAT GGATTTTTTT TTTGTAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 780

AAAAA

785

5

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 1069 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

TCCTCACCAT TCCCTAGGN CAGGTCCTG CAGGTCCTC ACTTCTCCCA GGTCCTAAA 60
CTTGGGTCGG TCCTTCCCT GGAGTAGCTG GNTCTCCAG TCGAGGTCCT TGTTCAGTCG 120
20 GTTCTTAGGC TCCTGCACAT GAAGGTGTGT GCCTGTGGTG TGTGGGCTGC TCTAGGAGCA 180
GATACAGGCT GGTATAGAGG ATGCAGAAAG GTAGGGCAGT ATGTTTAAGT CCAGACTTGG 240
25 CACATGGCTA GGGATACTGC TCACTAGCTG TGGAGGTCCT CAGGAGTGGG GAGAATGAGT 300
AGGAGGGCAG AAGCTTCCAT TTTTGTCTT CCTAAGACCC TGTATTGTGT GTTATTCTCT 360
GCCCTTCCGA GTCCTGCAGT GGGCTGCCCT GTACCTGAA CCTCATGAGC CTCTAAGGGA 420
30 AAGGAGGAAC AATTAGGACG TGGCAATGAG ACCTGGCAGG GCAGATACA AGCCAGCAC 480
CAGTGTCCCA GCCTTACTGG GTCCTTACCC TGGGCCAAC AGGGAGGGCT GATACCTCCT 540
35 TGCTCTTCT AGATGCCAC CTCTACAAT CTCAGCCAC AAGTCTCTC CACCCTAGGG 600
GGCTGTCTGC ATGGCAATAA CTCATAATCT GATTGGAGG TTTGCCCTTT ACAGGGGCAG 660
ATTTTCTGCT CAGTTCAACA ATGAAATGAA GAGGAACTCC CTCTTTCTAC AGCTCACTTC 720
40 TATCAGAGGC CCAGTGCCT CAGAGCCACA TTGAGTTGCT TTTTCTGGGA TGAGGAAGTA 780
GGGTAAACT CCCAGTTTC CTGAGGGAGG CTCCTGACAG GTGCCCTTTG TCAGACCCTA 840
45 CCACAGCCTG GATAGGCAGC CACATGGTC CTCGCCCTTG CTGGNACTC CGTGGTGGTC 900
CTGCCCTTCT CCTGCAATG CTGTGGGTCT GCTCTGGTGT GTGAAGGTCG GTGGGTAAAC 960
TGTGTGCCTA CTGAACCTGG CAAATAACA TCACCCTGCA AAGCCAAAAA AAAAAAAAAA 1020
50 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1069

55

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:

- 60 (A) LENGTH: 831 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

5 GGACATTAGA TCACTGTGGA CCTAAAACAA ACAACAACCT ATAAGGAAAA TGGCATTAGA 60
 AATGGTCTGG GGATCAGTTT ATCACTGCAG TTGTTACATC ACCCATGGT CTAAATACA 120
 10 GAGCTTTAGT CTGTCTCTGT TTCAGTTCAT TTTACAGGAG GTGAACATCA CACTTCCAGA 180
 AACTCTGTC TGGTATGAAA GGTATAAATT TGATATTCCT GTCTTTCCT TGAATGGCCA 240
 15 GTTCTGATG ATGCATCGAG TAAACACCTC AAAACTTGAA AACAGCTCC TGAAACTTGA 300
 GCAGCAAAGT ACTGGARGCT GACTGATGCC CTCATGATTT TCCACCTCT CTTCCTATAA 360
 AGCATCTTCC TAAGGAAATG AMCATGGCCT GATACTCAT TGTCACTTG TACAGAGCCC 420
 20 TAAGGATGTT CTGAATTCAG TGGTGCCAAA TAAATGTTGA CATTCCTCTT TTGGTTGATG 480
 GAAGTATCAG TGTGGGAACT GTTGTCTTAA TGGCATTTTA TAAATAAKA AKAKCATATT 540
 AGCAGGGAGG GAGATGATGG AGGAGGGAG AAGTCCATTT GTCTTATTTA TCCTTTTGT 600
 25 ATTAATAGAG AAGCACTTCA CAGTCACTGG CAATGCCATT TATAGGAAGA AGGTCTGCA 660
 TTCTGCTGC TCCCGGAGGG CTTAACTTTT TAATGAAAGA ATAAATGCTC TTCCACTCAG 720
 30 TAGATAAAGT GAAATGTGAA TTGTTAATAA CTGTGCACGG TCAATAAAGC GATGTTTTAA 780
 GGAATACAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAACTCG A 831

35

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:

40 (A) LENGTH: 590 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

TATATATAGA CNGTTAATAG TCGTGANTGN TGTGNACGAA CATTAACGGA AGTAGCATGT 60
 AGCCAGTCGA ATAACNTATA AGGACAAAGT GGAGTCCACG CGTCCGCCCG TCTAGACTAG 120
 50 TGGATCCCCC GGCTGCAGGA TTCGGCACGA GCTGCCAGGT GAGGAGCAGA GAGACTGTTT 180
 CCTTGGGTGG AGAGGTGTGG GCATGAGAGC CACCCATTGC CAAGCAGCAA GAATGTTTCGT 240
 55 GCTTTTTTCC CTTCCAAAAT ATGCAGGGCT CAGGCTCCCA ATTCCGGGCC TGTCTGCTTT 300
 GCTTGTTTCT CTCTGTCCC TGTCTCTCCG GAGGGCCAG GTGGAACCTA CGACAGGGAG 360
 GGAGACGCTT CCCAAAAACC TGCAGGGCTA TTTCOCAGAA TTTGGTTTTC AAGTACAAA 420
 60

	CTTTTGTGCC TGTAAGATAT ATGCAGCCTC ACAGAAGCAG CCTCTGCCTC CACTTTACCA	480
	GCTACGTTTT TATCTTAAGC ACATGGGGCT CCCTTAGAAC TTACTIONACT GATTAAAAA	540
5	AAAAAAAAA AACTCGAGG GGGGGCCCGG TACCCATTCT CCCTAAAAGT	590
10	(2) INFORMATION FOR SEQ ID NO: 77:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1274 base pairs	
	(B) TYPE: nucleic acid	
15	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:	
20	GAGCCACCAC ACCTGGCCTG GAAGGAACCT CTAAAAATCA GTTACGTCT TGTATTTTGT	60
	TCTGTGATGG AGGACACTGG AGAGAGTTGC TATTCCAGTC AATCATGTCT AGTCACTGGA	120
	CTCTGAAAAT CCTATTGGTT CCTTATTATT ATTGAGTTT AGAGTTCCCT TCTGGGTTTG	180
25	TATTATGTCT GGCAAATGAC CTGGGTATC ACTTTTCTC CAGGGTTAGA TCATAGATCT	240
	TGGAAGCTCC TTAGAGAGCA TTTTGCTCTT ACCAAGGATC AGATACTGGA GCCCCACATA	300
30	ATAGATTTC A TTCACTCTA GCCTACATAG AGCTTTCTGT TGCTGTCTCT TGCCATGCAC	360
	TGTGCGGTG ATTACACACT TGACAGTACC AGGAGACAAA TGACTIONTACAG ATCCCCCGAC	420
	ATGCCCTCTC CCCTTGGCAA GCTCAGTTGC CTTGATAGTA GCATGTTTCT GTTCTGATG	480
35	TACCTTTTTT CTCTTCTCTT TTGCATCAGC CAATTCCCAG AATTTCCTCA GGCAATTTGT	540
	AGAGGACCTT TTTGGGGTCC TATATGAGCC ATGTCTCTCA AGCTTTTAAA CTCCTTGCT	600
40	CTCCTACAAT ATTCACTACA TGACCACTGT CATCTTAGAA GGCTTCTGAA AAGAGGGGCA	660
	AGAGCCACTC TGCCCCACAA AGGTGGGGT CCATCTTCTC TCCGAGGTTG TGAAAGTTTT	720
	CAAATGTAC TAATAGGCTG GGGCCCTGAC TTGGCTGTGG GCTTTGGGAG GGGTAAGCTG	780
45	CTTTCTAGAT CTCTCCAGT GAGGCATGGA GGTGTTCTG AATTGTTCT ACCTCACAGG	840
	GATGTTGTGA GGCTTGAAAA GGTCAAAAA TGATGGCCCC TTGAGCTCTT TGTAAGAAAG	900
50	GTAGATGAAA TATCGGATGT AATCTGAAAA AAAGATAAAA TGTGACTTCC CTGCTCTGT	960
	GCAGCAGTCG GGCTGGATGC TCTGTGGCCT TTCTTGGGTC CTCATGCCAC CCCACAGCTC	1020
	CCAGGAACCT TGAAGCCAAT CTGGGGGACT TTCAGATGTT TGACAAAGAG GTACCAGGCA	1080
55	AACTTCTGCT TACACATGCC CTGAATGAAT TGCTAAATTT CAAAGGAAAT GGACCCTGCT	1140
	TTTAAGGATG TACAAAAGTA TGTCTGCATC GATGCTGTA CTGTAAATTT CTAATTTATC	1200
60	ACTGTACAAA GAAAACCCCT TGCTATTATA TTTGTATTA AAGGAAAATA AAGTTTGT	1260

TGTTAAAAAA AAAA

1274

5

(2) INFORMATION FOR SEQ ID NO: 78:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1133 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

AGGATTTTTC CTGTTCAC CAAAATCTGA GCATTCITT TATGTTGAAA AACTGAAAA 60
ACTAATTTWA GTTAATGAAC TAGAAAGAAT ATTGATTTTW AAGAAACAGA AAAATACTAC 120
TTATTTTCCT TCICAAATAA CGTTCTTTTC AAAAACCTCT GGCTGAAGTA TAACATGCTG 180
GTAGTAACA TAAATCTTGT CTTCTCTTG TTCTTTATCT TTCTTTGTTA TTTAGATGCT 240
TGTATAAATG TCTTTTGTIT TTAITAAGTG CCTAATGAC AGAGCTTAAT TTGAAGAAGT 300
GCCCTAATTT ATGACCACT TAAGAATTGC CTTATATGGG GTATTTTATT TGTCTCTGCG 360
TCTTTTGTAT GTGTTCAGT CTAATCATCC CTGTGAGTAT GTGTGGGGA CAGCTGATAG 420
AAGGGAGGAG AGTGTGCTA TGCTCAGGAT TGCCCTTTAG CCACTCAGCC AGAGATCCAC 480
AGGGAGCAAC AAGGACAGTT TCACATGCTT AGACTTTCTT GGAAGAAACA GTGAGGAGGA 540
GTAAGTCGTG AGTAGTGTA AGCTGGATGT AGAATGTCC TAAGGCAGTT GACCCACCT 600
TCCAACATGT TTTCACTTTA TTGCCCCCCT CCTACATTG GGTAGGTTT CATTGGATT 660
TGCAGCAATA ATGACTTTAT TTCTCTCTTG GTCAGGATTT GGCACATAAA ATCCTTTTAT 720
TATAGAACTA GCTATTTTAG TTACATAGTA ATGTAACATA TGGAGAGATT TATAGAGAAT 780
TTTGKTTTTG CTGTCATATA TGTCCATTTT GGAGACAGAT ATGATAGAAC TAGAAATTAA 840
GTTCATTTTC TGCAAGTCC ATTTGAATGA ACTTCAAGTA TCTTCTTAAT TATTAAATTT 900
TCTGATGAAG GCATTGTAAC AAATATATAG TATTATTAAA TCTAATTAAT ATTGGAAAT 960
ATTAATAAAT AGGTATTTTA TTTACTGTAA AAGTCAAAC TTCATTATGT AGATAAATCT 1020
TATTCPTTTC ATTCTTTCCC CTGTTTACAT CCTTTTACA AAGCTTAGTC ACCAATTAAA 1080
GCTTCTCTAT CAAAAA AAAA AAAA ACTGAGACT AGTTCTCTCT CCT 1133

55

(2) INFORMATION FOR SEQ ID NO: 79:

60

(i) SEQUENCE CHARACTERISTICS:

334

- (A) LENGTH: 661 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

GAATTCGGCA CGAGGGGAAA AGGATGCTGA ACGAGAGCAG AAAGCCTCTT TCCTTTGCTT 60
10 CACGCCCTTC CAGTCTTTAT TTAAACTCG GGTCCCTTT CTGTGGTCGC AGCAACCTTT 120
ACTCCACCTG CACTGCTGCT CCTGGGGGCT CCCAGGCCT CCTCTGCCT TTCTACCCAG 180
TGGCTGACGG GATGCCTGTC TTGCCTGGAC GCACCACTGC TCTCCTGTCC CTCACCTTGG 240
15 CTTTGTGCTGT GCCCTGCTCT GGGGTGAAG CTGGCCCATG TGTCCCCCG AGTCATGGCT 300
GCTCCTCCTG GGAGGCTCT GTGTGCGTCA CGTCTTCCAC ACCTGGGGG AGCTGGCGAG 360
20 CCCGTGCTCT GTTCCCTCG GCTGCTTGGC ACAGAGTGC AGCCTGGGAY TCTCCGTGGA 420
CCCAGACTGG GGATTTTGCC AGGGGGGCGA TGGGAGGAGC AGGTGCTTTG CCTGGCGGCT 480
GTGTCTGCAT TTCTGGACGC CCCAGAGCAC AGAAGTTGCC GGCACTTTGA GGTCTTCCTC 540
25 GGCAITGCC AGATTACATG AGTGACGGCT GGAATATGT TTTCTTTTTT GTAATGGAGG 600
CGTGTTCAC ATATAGTAAA GCTACCAAA AAGTAAAAA AAAAAAAA AAAAACTCG 660
30 A 661

35 (2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1378 base pairs
(B) TYPE: nucleic acid
40 (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

45 ATTGGGTACC GGGCCCCCCC TCGAAGTTTT TTTTTTTTTT TTTTAATGAA AGCTCTCAAA 60
TAAGCGATTT TATTCCTATC CATGATTGCA GACATTTACA AAACCATAAC ATCTGAGTTC 120
ACCTTAAAAA ATAACCTATA TAAAGCAGTG ATATACACAG CACAAAATAG TTCAGGGAGG 180
50 GGGCAGGAGC AACTTGTAAT AATTAAAATG TAAACGTGAA AAAAAGGATG GAATAAAAGT 240
CCCTACTTAT TTCTACTTAA GATGTCATGT GATAATATTT TACAATGTCC TGTGGGTCAA 300
TGTATGTATG TGTATATGTC TGTATAACAT ACACATATAC AGTACATTCT CTTTCCCACA 360
55 CATATACATA CACACATAAT TATTTGCAGT TCAGTTTAGG GCAATTCTAA TATGCCACTC 420
CGTACAGTTG TTTGAATCAC ATTTGGACCC GCTTCTTCA CAAAAGAGGG GAGAGAGCAG 480
60

5 GAAATAAAAA GGTTGGTTTG GTGTGACTGA GATTCCCTTG TTAACTGTA CACTGTGATG 540
 AATAATTTTC TTCCGTAGTA GTTCTGTGAA GGGCTGACTC ACTGTGGTTT TCATGAGGAG 600
 10 ACTTGGTAAT GGATCACACG CTCATTGTCA TGCTAGGGA GTAACCTCA CTCTGAAAAG 660
 GATTTAAGAA ATTTCCCCC ATTTGCCAT CATCCCTTGG AGTGCCCGT TGATTACTCA 720
 GGCTCATATT ATGGGAGAA TTCTTGAAA TACTGTCCAT ATCTCTGAG CCTAAAGAGC 780
 15 CATTCATGTG ATGTGACTCC ATTCTCCTA ATCCACCCAT GGGACCATCT GACCCAGGRC 840
 CCATTGAAA ATTAGTCTG TTAGGTCCAG GAGGTACTGC ATTCATTAAA GTATACATGT 900
 TATCACCAGA GTTGGTTGAA TCTGCTGGAC TAGGCATGAT GGGTGTCTCT GGTGGCCCTC 960
 CACCTCCTGG AGGACCTACA TAATCCCAG GAGATGCTGA GGAGTATGGT ATTGAATTGG 1020
 20 CATTTGTTGG GTTGGCCAA GGTCTACCAC CACCTGGACC CATGTTTATT CCAGGCATTC 1080
 CAGGGCCACC TAAAGCATTC AGTGGGGTC TCATTGCACC TCCATAGTTC TGTGGTCTTA 1140
 AGGGCACCAT TCCTCTTGA GGAGTCATTC TCTGCATTGG CCCACCCATA TTTGGATGTC 1200
 25 CTTGTGTGCG AGTTGGATCC ATTCCACTGG GGAGTAATGG CTGACTTCCT GGGACACCTC 1260
 CAAGTGCTG ATTAGGTATC CTCAATGGG GCCTTGGACC TCCAGGGTAC CGAGGTGACA 1320
 30 TAAAAGGTA ATCATGGAAG GCTTTTGTCT CACTTGAGTG TTCACATGTT TCACGTCT 1378

(2) INFORMATION FOR SEQ ID NO: 81:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1440 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 40 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

45 ACTTTGTCCA AATGTGCTG TCACATGTAG TCAGCTGNAG NAATTAAAA TGAATTGCCA 60
 AGTGAAGAGT CTGTGGATTA ATTGGCGTT AATTAACAGG CTTTATCAAT GTGTCTCAA 120
 GGGAGAGGCC CAACCTAAT TAAGGAGCTA AACTTCCTGA GTGAGGGCT GTGAGGATGG 180
 50 AGGTGGAGGA GGCATCTGG GGGGTGGTG GCCGGCCAG CAGATGGCGC CTCCTGGCT 240
 GAGCTGCCCG CACCGCCAGT TCCCTCATTT CCACTCAGGA AGGCAGAGAA GGCAGAGTGA 300
 TCTCTCAAG GAAGAGCTTC CCCAGCCTC GGGAGCAGCT GGCAGGGCT CCGGAATAA 360
 55 GCCCTACAG CCGCGCCTG CCTCCAATC ACTAACCCTG CGCCTCTGT CTTTCAGATT 420
 CAACGCGTTC AACAGAAGCC ATCCCAGCC CAGCTTAAAT TATAAGATA GACAATAACT 480
 60 CTGTTCAT CTGCGTGGTG CTCTTTAGT AAATACTGTA CAGATTTAC CATGGAGAAC 540

TTTT TTTT TA GTTTT TACCT TTTCTTAATT ACCCTTATTC CGAATGGACG AACACTTTCT 600
 5 ACCACTGCTG ACCATTGTAA AATACCGTGT ATATAAATCC CATTGAAATA ATGCCCTGGA 660
 ATAGAACATC TCAAAATGCTG CTTAATTACA GACTCAGGTC GATTACTTGT ATTTTCATGTA 720
 ATGTTCTCTC AAGTTAGACA TCTGGTGCAA GACCAACCGG GAGACCATGG AATTGTCAAA 780
 10 AGTACAAACT GACAGTGTGT ATATTTAATT TAAAGACTTA TTTAAAACT CACAAGCTCT 840
 CACCTAGACT TTGGAGAGCA GTCTGTTTTC TGTAATGTCT GATACTAGAA ACTAATTTGC 900
 TTAATTTAGT TGTATTCAAG ATTTGAAGAT GTATTTTATA GACAAGTTCT GTTTTGAAC 960
 15 TTTGTGGAAC TGTTC AATC AATCAATTC CCAGTTATGA TGAGTATTTA CATTATGAAT 1020
 GTATAACCCA GACATGATTT GTAAAGCCGA CAGTATGTTT CTATTACACA ACACTTTTGT 1080
 20 ATACAGCGTC TCTGTCTTC ACTGATACTG GAGTCTCCGT TGCTGCGNNG GTCCCTTCGA 1140
 GTTCTAGTT ACAGACACAA TCATACTGTG ATTTTATTTT TAATATGGAT ATGCTATCAA 1200
 ACTGTGATAC ACTTATAATT CACTGGTCTT GCATCAGGAG ATGGAGTGGG GAAACTGTA 1260
 25 TTTAATACAG TTTGTATCTG AATAATCTGT ATGGTTTATA CAGTTTGTGT TGTTCAGAGA 1320
 TGTTTAAAGT TTGATCTTTG TTTTCTTAAA GATTAAAAAA GCACTTGCCC CACTGTAAAT 1380
 30 ATACAGCATG TAAATTTCT RTAGTATATA AATGGCAGCA AATCACA AAAA AAAAAAN 1440

35 (2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:

40 (A) LENGTH: 1381 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

45 CCGGGGCTGC AGGAATTCGK YACGAGGCCA GCAGTTGCTC CCAGTTCAGG AGGTGCTCCT 60
 GTACCCCTGGC CACAGCCCAA TCCTGCCACT GCTGACATCT GGGGAGACTT TACCAAATCT 120
 ACAGGATCAA CTTCCAGCCA GACCCAGCCA GGCACAGGCT GGTCCAGTT CTGACCTGAG 180
 50 CACGGTTTTT CCTCATGTGA CTCTGGGAA GCGCTCCCT CATCTGGGCC AAAGGAAGGA 240
 GGACGAAGCC CTCCTCAGCT GGCTGTGTT TGGGGCATGA ATCTCTCTC TCCTCCTTGT 300
 55 CTGGCTCTGT TGACAAACCG GGCATGTTG GCAGTAAAT GGCACCGTGT CACACTGTTT 360
 CCTGGGATTC AAGTATGCAA CCAGAACACA GGAGAAGAAA AGCTCCAGGA TCCCTGTCCC 420
 CATCTGTCTT CTGATGTGA GAGAGACTCT GAGACTTCTT CCATCGCAAT GACCTGTATT 480
 60

	AAACACAAGC CCCCCAAGCA AAAGAAGAGG TTGAGTTTGC TGCCAGGATT CAGATCAGCC	540
	CTTCCCAGGG TCTGCAGGTG TCACATGATC ACAGTTCAGC GGGAGGCTTT CCGTACCCAC	600
5	ACTGGCTGTA GCACTTCAGT CCATCTGCCC TCCAGAGGAG GGTTCCTTCC TGATTTTATG	660
	CAGGTTTAGA GGCTGCAGCT TGAGCTACAA TCAGGAGGGA AATTGGAAGG ATTAGCAGCT	720
10	TTTAAAAATG TTTAAATATT TTGCTTTGCT AATGTGCTGA TCCGCACTAA CTCATCTTTG	780
	CAAAAGGAAC TGCTCCCTCG GCGTGCCCCA GCTGGGGCCT CTGAAGGGAT TCCTCACTGT	840
	GGGCAGCTGC CCTGAGCTTC AGGCAGCAGT GTTCATCTCT GGCCAGTTGT CTGGTTTCCA	900
15	TGTATTCTAG GCCAGGTAGG CAACACAGAG CCAAGGCGGG TGCTGGAAGC CAGACGGAAC	960
	AGTGTGGGG CAGGAAGGTG GATGCTGTTG TCATGGAGCT GTGGGAGTTG GCACTCTGTC	1020
20	TGCTGGTGGC CCTCTCGGCT CACATGTTCA CAGTGCAGCT CCTGGCAGAC TTGGGTTTTC	1080
	TCTTTGGTGG TTTCTAAAGT GCCTTATCTG CAAACAACCT CTTTCTCCT TCAGGAACG	1140
	TGAATGGCTA GAAGAAGGAG CTCAGTAAAC TAGAAGTCCA GGGTTGCTTG GTTACTGGT	1200
25	TTATAAGAAA TCTGAAAGCA CCTCTGACAT TCCTTTTATT AACTCACCTC TCAGTTGAAA	1260
	GATTTCTTCT TTGAAAGGTC AAGACCGTGA ACTGAAAAAA GTGTTGGCCT TTTTGGCGGA	1320
30	CCAGATTTTT AAGATAAAAT AAATATTTTT ACTTCTGTCA AAAAAAAAAA AAAAAAATNT	1380
	C	1381

35

(2) INFORMATION FOR SEQ ID NO: 83:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1706 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

40

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

45

	ACTGCACCAC TGCCAGGTC TCCCGGCTGG ATGAAGACGT GGTCCATGAG GAAGCTGGCT	60
	AGCTCAGACT GGAGAGTAGC TTCAGGAAAA AAGACAAGTG GCCTAAGGAA ATCAGGCCC	120
50	CCAACTATCA TCTGAGGGCT AAAGATGAGA AGTAGATCAC TTAATAAGAC AAAAGCCTGT	180
	AGGGGAAAA GAAAGGATGT TTAAAAGGAC AGAATGTTTC CCAAGGTAGA AATGACACTG	240
55	TCAATTTCTC CTGGAATGG GGCAGGGAT ACTCGCCTTG TTGCTCCAC TTGAGTCAGT	300
	ACTCACCTGC TCCTGGATCT CAGTATCCAC ATCTGAGAGG CAACTCTGGC AGAGTTCACA	360
	GAAGGCCACC ATTCTGTCCC TCAAACTCGA CAGCTGCTTC TGTGGGCACA GTGGCTTGAA	420
60	GGGAAGAAT GAAGACACAG ACTCCTCTGT TOCCATTATC CCATCTAAGA CCCACACTCA	480

5
10
15
20
25
30
35
40
45

CCTGGGAAG CATCTGATTT AGAAATGTGG GTTAGTGTCC AGAGAATGGA AAAATAGACA 540
AGAGTCAAGG CTGGCAGGAT AACCTGTAAC AACAAAGGGT TTGAAAAATG AGGTTTGGGT 600
TAGGAGAGGG AGAGACAGAT AGCCAGAAAC ACACCACTGA AGAGGAGAGA AAATGAGTAA 660
AGGGAGAGCT AATTCCTTTT CCAGTGGAAA ATGAGTGATA TTCTGGACAT TCTTCAGAGG 720
CATCTACACG AAGTAGAAAT GTCACCGCTC CCTAATTTC TCTACGTCCT CTAGAATCCC 780
TCAATATTAT CCTTGGCTTC CAGGAAATCC AAGAAGACCC TGGAAGTAGA GTCCACCTTC 840
TAAGAGAGGA ATGTAAGAGG TGACCCCCAC CCACCTGATC TTCTCGCTT TGTCCACTCC 900
ACGCACTGAG ACTTGACACA CCTAGTGGCC ACCTAGAACG TAGGTCCTTA AAATYTAGCC 960
CCCCAGCCCC CAACCCATCT CTAGCCTGTC CACTCACCTG GTGAGGAACY TYTCTGTGT 1020
CCACAGCYTT CTGCAGGAGT TGGCAACATG GCTCATAGAG CTCCCAGCGA GTCAGGTCAT 1080
GAGTGCTTTG GGGAGAGAAAG GGAATGTTA TACTGGAAAA GAACAGAGGG AACCAACTCC 1140
ACAGACACCA GTAAAAACGG GATGGGAAG AGGAGGAAAG CCACTCACTT GTAGAAGGCA 1200
GAGAGGCGTT TCAGAGTGGC TGCCAGATTA TATACCTCAT CCTCATCTAG GAAGGACGAC 1260
TGAGAAGGAA AGAAGATCCA CAATAGCATT TCCCCAGAA CTCATCAGTC CACATCCCCC 1320
GTCTGCGAGC CCCTCCACC CTGTGTTGGG GTGTCCCATT GTCCAGCCCC AGCTCCTACC 1380
TGTAACAGCT CTTCAAGCTC CTGCTGGAAR CGTCAGTCA GCAATCTAC TAGCTGGCTG 1440
CGGGCAAAGT CCGCCCGCT GAAGAAAGTG AATTCGGGAT TACAGAGCAG GTAAGAGCAT 1500
GGCCCCAGC CTCAAGCACC GCTGGCTCTG CATGCTTCAC CACCACCTCC TGGAGTTGCT 1560
GCAGGAACAG CTCCAGGTGC TGAGAAGAAA AGGCAGAAGA TGGTGTGCTG TGGGGATGGG 1620
AGGAGGACAC TCTTCTGGCG GGAAGTGAA CGGGTTAAA AGCATTAAAC TTCAAGGATA 1680
AGATGCCTAA RAAAAAAAAA AAAAAA 1706

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 573 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

GAATTCGGCA CGAGCTTGGT AGCCTTAGAA CTGCATGAGC TGCTTTACCA CTGGGAAACA 60
CGAGCACAGC CTAGCTTGAT TTTGTATGTG GTATCAGATC TAAGGTGGAT GGAATTCAGG 120

ACTTCTCTGTC TACTCTTTGA TTTTGTTTTA TTTTGTAGAA TGTTTATTT TGTTTATTC 180
ATTTATTCAT CTTAGAGAC ATGGTCTGGC TCTGTGCCC AGGATGGAGT GCATGGTGTG 240
5 ATCATAGGCC ACTGCAGTGT TGAGCTCCCG GGCTCAGGCG ATCTCTCTGC CTCAGCTYCC 300
TTAGTAGCTG GGACTATAGG CACATGCCCT ACCATGCCCTG GCTTGTCTA CTTTTGAAT 360
GATGTCYCAA ACTAGAAGGT CTATTAATTT AAAAAATTAA GGATAGCATG CCATAATTAA 420
10 AATAATAAC AGTGGGAAAA GGCACCTTCC AATGATTCAG ACATCAACTT GTGATTAAA 480
AAAACGAAAA ATAAATAATA GGAAAAAAG GGGAAAAAGT TAAATAAAAA TAAAATTAAA 540
15 AAAAAAAAAA AAAAAGCTGA GGGGGGCCCC GTA 573

20 (2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 684 base pairs
(B) TYPE: nucleic acid
25 (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

30 CTCITTGGCT GTGTCTACCT CCTTCATCTG CTGGCCGAC ATAAGCACCG CCCTGCCCT 60
AGGCTCCAGC CGTCCCGCAC CAGCCCCAG GCACGAGAG CACGAGCATG GGCACCAAGC 120
CAGGCCTCCC AGGCTGTCT YCAGTCCCT TATGCCACTA TCAACACCAG CTGCGCCCA 180
35 GCTACTTTGG ACACAGCTCA CCCCCATGGG GGGCCGTCCT GGTGGGCGTC ACTCCCCACC 240
CACGCTGCAC ACCGCCCCCA GGGCCCTGCC GCTGGGCTT CCACACCAT CCCTGCACGT 300
40 GGCAGCTTTG TCTCTGTGA GAATGGACTC TACGCTCAGG CAGGGGAGAR GCCTCCTCAC 360
ACTGGTCCCG GCTCACTCT TTTCCCTGAC CTTGGGGG CAGGGGCAT GGAAGGACCC 420
TTAGGAGTTC GATGAGAGAG ACCATGAGGC CACTGGGCTT TCCCCCTCCC AGGCCTCCTG 480
45 GGTGTCATCC CTTACTTTA ATTCTTGGG CTTCAATAAG TGTCCCATAG GTGTCTGGCC 540
AGGCCACCT GCTGCGGATG TGTCTGTGT GGTGTGTGG GCACAGGTGT GAGTGTGTGA 600
50 GTGACAGTTA CCCAATTCA GTCAATTCTT GCTGCAACTA AGTCAGCAAC ACAGTTTCTC 660
TGAAAAAAAA AAAAAAAAAA AAAC 684

55

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1036 base pairs

60

340

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

TGGAGGCAGA TGCACAGGAG AAAGGTTCCT GTCCGCACCC TCTCAGACCT GAGGCTGAGC 60
TTGCAGTGAG GGCTTCTCCT CGGCCCTCG CCGCCCCCA GAGTGCCAT CCTGTCTGTT 120
10 ACAAGCCAGA GGAGCCCGGA TGTGAGGCCC CAGATCACCT CCAGGGACTT GGGGTTCCCA 180
TCTGAAATCC TTTATTTTGT TACCATGGGG TGGGCCCCGG GCTGAGAAGG AAGAAGCACC 240
15 CTCTCCCCGG CCTCCTCTGT CTGCACCCGT GGGGCTGTGA CTTACTCCTG CCTCCAGGGG 300
CGGGGCGGGG CCCCTGGGA CCTCTTAAGG CCAAGGTGG GCCCCAGGAC CTYTGGCAG 360
AGTGGAYTGC TCATGGCAGA TGTGTGGCAA TGCTGGCTG WGTCTTTCCG GCAMCTGCGT 420
20 YCCCTYTCCC GGGYTCCCT GCTGCATGGT GGATGTGCTC CTTCTGGCC CGGTACATT 480
GCCTCCTTGA GCCTTAGTCC AGGGGGTCAC TYCTCCACC CCACCTACCT CACAGGGTTG 540
25 TTGTGAGGT GCACAGAGGA GCAAAGTCCC TGAAGGCCCT CAGGCAGTAT ATAGGGGCCG 600
CCCACCTTCA GCTGCCCTGG GATGGGAAGG ACCCAGCCCG ACCCTGGGC ATAACACTGT 660
GTTTGCAAT GGAGATTCAG GTATTGGGA TGCAGGTGT GGGAGCTGG CCTGGCAGAG 720
30 TAGGGTAGT TGGCTTGGC TTCTCTTGG TGATCCACC CCCAGCCATT TGCATTGCTG 780
GCCCAGCGCC TGGCTGGG GCGGGGAGA GGCAGCAGAA GGGCTGGC AGGGGCGGTG 840
35 GAGGACTCAG GAACTGCCG GGGAGAGTGG GTATGGCGC TGAGCCAGG GCCCTCCTGT 900
GTTTGACTTC CCGGATGGG TCCTTGCTTC TCAGCTGTGT CCGACCCAC CATGTAATAA 960
AACCACAAAG AACAGCAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAAN 1020
40 CCCNGGGGGG GNCCCG 1036

45

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:

50 (A) LENGTH: 908 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

55 TTAACAAAT GGAATCATGC AATATGTGAC CTTTTCGTC TGGCTTATTT TATTTAGCAT 60
AATGTTTTTG AGGTTTCATCC AAGCTGTAGC ATGTATCAGC ACCTCATTTT TTTTCTGGC 120
60 TGAATATTAT TCCATTATAT GGATTTACCA CAATTCATTT ACCTATTCAT CTTTGTGTTT 180

5
 10
 15
 20
 25

TGCTGTCTGG CTATTGTGAA TAATGCTTCG ATAAACATTC ATATACAAGT TTCTATGTGG 240
 CTTTATGTTT TCATTCTCT TGGCTATCTA CATGGGAGTA GAATTCTAGG TCATAATATA 300
 ATTTTATGTT TAACTTCTCA AAGAATGCC AAAAGGTTTT TCATAGTGGC TGCATCATTT 360
 ACATTCCAC CGGCAATGTA CAAGGATTTT TATTTTTCCA TATCCTTGCA CTTACCAACA 420
 CTTCTTTTTK GIWATWATTT TGTTTTTICA TTATIGCCAC CCTAGTGGAT GTGAAATGGC 480
 ATCTTATGTT TTTGATTTGC ATTTCTCTAA TGACAAATGA TATCATACTT TTTTATGTG 540
 CTTACGGATC AAAGGTATTT CCTTGGAGAA ATGTCCCTTC AAGTCCTTTG CCATTTCAAA 600
 ATTTGGTTAT TTGTCTTTTA TTATTCAGTT TTAAGAAATT CTGGCCAGGC GCAGTGGCTC 660
 ACCTGTAATC MTAGCACTTT GGGAGGCCAA GCGGGCAGA TCACCTGAGK TCAGGACTTC 720
 GAGACCAGCC TGGCCAACAT GGTGAAACCC CATCTTACTA AAAATACAAA AATTAGCTGG 780
 GCGTGGTGGC AGGTGCATGT AATCNTATCT ACTCAGGAGG CTGAGGCAGG AGAATCGCTT 840
 GAACCCAGGA GCGGAGGCT GCAGTGAGCC AAGATCACGC CATTGCACTC TAGCCTGGGT 900
 GACACAGA 908

30

(2) INFORMATION FOR SEQ ID NO: 88:

35
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 655 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

45
 50
 55
 60

TGCACTGGTT CCTTCTCCCC AGCAAATACT GCCTTCTTGT TTTTCTCTGA TGTGGCAGGT 60
 GACTACAAAA TCCGCCTTGG TATTCTTCAA ATGCATATAT ATTCTTTTCT TGTCACTCC 120
 CTCTCTTCTT AGATTAGAAA ACTGCCTCAT TTCTGTCTCA CTGGATGTGC AGTCCCAGCT 180
 TGTCTTCTCT TCTTCCCCCT CTGTTCAGG TGTCTTTTTT TTTTCTCTTC TCTCCCCACT 240
 GGGCAGCAAA AGTTGTTCCA CAGTGGAAAW TTAGGCATCC TCAAGTTTCY TCCCAGCTTC 300
 TGCTGTGTTT TCTTAGAGTA AATGCCAAT TTCTGTTTTT ACAGGAAATC CTTTTTTAAA 360
 AATGGAATCA GTGTGGTCCC CATCTACTCT GCAAAAATG CATTTTTCTC TATTTTCAAA 420
 TGAGATTTGT TCAAGTTTCA AAACCACGTG AAATAATAAA TGTATAGTAG TTTTCTTTTC 480
 CTTGGGCATT GCTWGATATG TGAAATGGGT TTATGAAAAA TAATAAAATC ATAACGCTAT 540
 TTGTTTGACT TTCAATTCA TGGGAATTTT TCTCAGCTAA ACTCTAAATG GTGATTARGC 600

AAAAAAAAA AAAAAAACY GRAGGGGGGC CCGGTACCAA TTCGCCCTAT AATGA

655

5

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 1102 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

15

TTTTTTTTT ACCATTAAAA ATAAATGAA AGTGACCTTC TGTTTATAAA AATCTTTGTC 60

TGCATCTCTG CTTATTTCCT TAGAAGAGAT TCCAAGAAGC GGTGAGTGAT TTCACGGCAG 120

20

CAGAGGGTTG GGACATATTA CGGGCGCGGA TCCCTCTTGG AGTGAGATGA CTCTCCGGAG 180

AGATTAGTC GTCACCTCG CGTGTGAGGC TCGTCACAC CCCAGGGATG TGTCTATCAA 240

GATGGAAGAT CTTTACACG CTCTTGATT TGTGTGCTY TTTTCTATT ACTAGTGAGA 300

25

AKGAAACTTT TTATATGATT ATTATCCATC ATAATCCAAC ACAATTACT GCTTCATGTT 360

CTTTACTTT CCGTGAAGG TTTTAGTGCC TTTTAAAAAT TGCTATATAT TAAGCTTGTT 420

30

AATACTCCA TGCCTATTT GTGGSCATCA RPTCCCGG GNACAGGCNT GCACATTTG 480

CCTTCACAG CTGGGTGGTT TTTCAITTT AMTCTATTT CTCGTCTTC TATCGTTTA 540

TGTTACAGAC GGTTCCTCG TGTAGAAAGC AGTTTATGAA GATTFACTTT CGACAGTCTT 600

35

CTCTCTACTT TCTACAGTA ATTCTCTGAT GTGTCTGGGA GTTTGGGGT CTGGGTAAGA 660

RTCTCTCTT CACCTATTC TCTATTACGA TCCACAGCCT CATGCTTTAT GARATTGGTG 720

40

GCGGGGARG GGGGAGATT GCGGATCCCC CAAGCCAGAC TTTATCCCC TATCCCTGCC 780

TCTGGATCCC ACGTACAGC CTGGGAATC CCGTGGGTA GGGGCAATG GTCTCGCACT 840

CTCACCTGTA CCCCAGGGCT GGCACAGGAT GGTCAAGGAG AGAGGCTGCC CAAGCGCATC 900

45

CYTCTGGTGT CCCCCTGACA CGCTCCAAA GTGAGCAGGT AGGTTTCAAC AGCCCCACGT 960

TGCAGGTGGG AGATGAAGCT CAGGGTGGAG ACCAGTATCT CACAGTCTC TTTGCATGCC 1020

50

CGGGTACTTG TTAGTCAACT GATCAAGTGA AAATCTAGC CCCAGAGGCA GGAGAATCCG 1080

GAACAAAATT AAACCAGCCA GG 1102

55

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:

60

(A) LENGTH: 1533 base pairs

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

	GGCACGAGCC GNCACGGGCA GCGCCCCATA GCGCCAGGA CCCCTGGCA GCGGGAGCCG	60
	CGGGTCGAGG TTATGGATCC AGCGGGCGGC CCCGGGGGCG TGCTCCCGCG GCCCTGCCGG	120
10	TGNCCTGGTGC TGCTGAACCC GCGCGCGGC AAGGGCAAGG CCTTGCAGCT CTCCCGGAGT	180
	CACGTGCAGC CCCTTTTGCC TGAGGCTGAA ATCTCCTTCA CGCTGATGCT CACTGAGCGG	240
15	CGGAACCACG CGCGGGARCT GGTGCGGTGC GAGGAGCTGG GCCGCTGGRA CGCTCTGGTG	300
	GTCATGTYTG GAGACGGGCT GATGCACGAG GTGGTGAACG GCCTTCATGG AGCGGCCTGA	360
	CTGGGAGACC GCCATCCAGA AGCCCCGTGT TAGCCTCCCA GCAGGCTCTG GCAACGCSCT	420
20	GGCAGCTTCC TTRAACCATT ATGCTGGCTA TRAGCAGGTC ACCAATGAAG ACCTCCTGAC	480
	CAACTGCACG CTATTGCTGT GCGCGCGGCT GCTGTACACC ATGAACCTGC TGTCTCTGCA	540
25	CACGGCTTCG GGGCTGCGCC TCTTCTCTGT GCTCAGCCTG GCCTGGGGCT TCATTGCTGA	600
	TGTGGACCTA GAGAGTGAGA AGTATCGGCG TCTGGGGGAG ATGCGCTTCA CTCTGGGCAC	660
	CTTCTGCGT CTGGCAGCCC TGCGCACCTA CCGGGGCGA CTGGCCTACC TCCTGTAGG	720
30	AAGAGTGGGT TCCAAGACAC CTGCCTCCCC CGTTGTGGTC CAGCAGGGCC CGGTAGATGC	780
	ACACCTTGTG CCACTGGAGG AGCCAGTGCC CTCTCACTGG ACAGTGGTGC CCGACGAGGA	840
35	CTTTGTGCTA GTCCTGGCAC TGCTGCACTC GCACCTGGGC AGTGAGATGT TTGCTGCACC	900
	CATGGGCGCG TGTGCAGCTG GCGTCATGCA TCTGTTCTAC GTGCGGGCGG GAGTGTCTCG	960
	TGCCATGCTG CTGCGCCTCT TCCTGGCCAT GGAGAAGGCG AGGCATATGG AGTATGAATG	1020
40	CCCTACTTGT GTATATGTGC CGTGGTGC CTTCGCTTG GAGCCCAAGG ATGGGAAAGG	1080
	TGTGTTTGCA GTGGATGGG AATTGATGGT TAGCGAGGCC GTGCAGGGCC AGGTGCACCC	1140
45	AAACTACTTC TGGATGGTCA GCGGTTGCGT GGAGCCCCCG CCCAGCTGGA AGCCCCAGCA	1200
	GATGCCACCG CCAGAAGAGC CCTTATGACC CCTGGGCGCG GCTGTGCCCTT AGTGTCTACT	1260
	TGCAGGACCC TTCTCTCTTC CCTAGGGCTG CAGGGCCTGT CCACAGCTCC TGTGGGGGTG	1320
50	GAGGAGACTC CTCTGGAGAA GGGTGAGAAG GTGGAGGCTA TGCTTTGGGG GGACAGGCCA	1380
	GAATGAAGTC CTGGGTCAGG AGCCAGCTG GCTGGGCCCA GCTGCCTATG TAAGGCCTTC	1440
55	TAGTTTGTTC TGAGACCCCC ACCCCACGAA CCAAATCCAA ATAAAGTGAC ATTCCCAAAA	1500
	AAAAAAAAA AAAAAAAAAA ANCCCGNGGG GGG	1533

(2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 575 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

ATCCTCTGGA ATCTAGGTGG AAGCCACCAA GCCTTCTTCA CACTTGCGTT CTGAGCATCT 60
 GCAGACTTAA CCCCATGTGG CAATCACCAA GGCTTATGGC TTGTGTCTC CAGAACTGTG 120
 15 GCCAGAGCTG TACCTGGGCC CCTTTGAGCT GAGGCTGAAG CCAGAGTCTG AAGCTCAGCA 180
 GGGCAGTARG GGCCTGGGCC TGGCCCTGA AACCATCTT TTCTCCTAAG CCTCTGGGCC 240
 20 TTTGATGGGA RGGCTGTCC TCAAGATTTT TGAAATGCCT TTGGAGGGTT TTTGCCTTGT 300
 CTGGATATT GGCTTCTTT TAGTTATGCT CATCTCTTA GCAAGTGAAT GTTTCACAAC 360
 CTGCTTGGAT TCTTTCTTA CCACAGARCC AGGCTGCAA TTTTACAAAC TTTTCACTC 420
 25 TGTTCCTT TTAATATAA ATTTCAATGT TAAGTCACTT CTTTGCTCCC ATATCTGATT 480
 TAGGTGCTG GAAGTAGCCA AGTCACCTCT TGAATGCTT GCTGCTTAGA AATTTCCTCT 540
 30 ACTAGGTAGC CTGGGTATC ACACCTAAGT TCAAA 575

35 (2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 639 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

45 TCCTTTCATC TTAAGCACCA CCCGACAGG CAGGTACTAT TACCATCTCC GTTGACAGA 60
 TNAGGAACCT GGCACAGGAA GCATTTAAGT GGATTCCCCA GGATCGCCCC ACTGTCAGGA 120
 GCAGATCAG AATGGGCTC AGCATCAGG TCCCAATCCT GGCTTCTAAC TGCTGCGCTC 180
 50 TGCCCTTCYC TCWCCCCACC TCCCCACTCC AGTGCCTTTG GTCATGCCAC TGCAGCTTTC 240
 AGGCCAATAC TGGATTAGC TCTTAGTGTT CTTGTCCCTG CAGCCATTTC CCCAGGCAGC 300
 55 AATTCCATGT GCCCTCACTG ATGTAGGTGG CTCTTGTTG ATTTGTACA TCCTATTGAA 360
 TTGTTTATGC ATCTTGTTCA CACTCACAGC ACCCTCCCTC TCACAGTCC TCCTTATAAA 420
 AATGTCCTC AGTGTCTGCT ATGAGCCAGG TGCAGACTTA AGTGACAGG CTGCTACGGG 480
 60

345

AAATAAAAAA TTAACAAGGA GCACCTGCCT CTTAATGCAC AGTAACAAAC TATGTTAAGT 540
GTCAGGAAGG AAAGGTTAAG GATGCCAGGA AGGCTTTTAA TAAATAACCT GACTTAGATG 600
5 GGCAGGTGGT GCTGARGATT AAGAACGTGT TCTTCTCGA 639

10 (2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 744 base pairs
(B) TYPE: nucleic acid
15 (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

20 GAATTCGGCA CGAGAGTGGC TGGAGTCTGG CTGCAGAGGG AAGACATCAG CAGGGAGGGA 60
GCCAGGGCCT GTCACATCTT TCCTCTGGCC ATTGTCTTGG TCTTTGTAAG CCCAGAATCT 120
CCCCCTCCCT GAAGGGAGGC CAGCACCCCA GGAGGGCAGC AGGTGTGCTG TGAGGGTTGG 180
25 AGTAGTGTGA GAGGTCAGGG TACACTAGAA TGGCCATGGA CACCATGTGG GGGTGTCTTG 240
GGCTGGGCCA CAGAACAGTG TCCTTCTCTG TGCTCTCTCC CTGCAGCTTC CCCCGACCTT 300
30 GTNGTTTATT TGGTTTGATA CCAATCAGCA GACCCTGCAA GGTGGAAGCT CCCAGGCTCT 360
CAGTCCACCS ACTCTCATGT GCCAGTCACC CTTACTGTAA CTGCCCAATG AGTACTTCTT 420
GCCCACTGCC AAGATAGAGC CAGTTTACCA AGACAGGGGA ATTGCAGTAG AGAAAGAGTT 480
35 GAATATACAT AGAGCCAGCT AAATGGGAGA GTGGAGTTTT CTTATTACTT AAATCAGCCT 540
CCCYTAAAT TCAGAGGTGA GAATTTTTC AAGACAGTTT GGTGGSCAGG CCTAGGGAAT 600
40 GGATGCTGCT GATTGGCTAG GGATGCAATC ATAGGGGTGT AGAAAAGTWC CTGTGCACT 660
GAGTCCACTT TTGGTGAGAG CTACCAAGGA GCTGCTGGTC TGCTGGTCCC GGTAGAGCCA 720
TCTGGTGTCA GGAATGCAAA AGTG 744
45

50 (2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 526 base pairs
(B) TYPE: nucleic acid
55 (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

GCAGGGGAAT TCGGCCACGG AGGGGTTTCA ACAGGGCCCG TGGGGTGAGG TGCAACACA 60
60

346

AAGCCCATAA GTGCTGGCCT GTTGGGACAA ATGAGAGAAA TCCCATAGGG TGGTGATGAC 120
 AGCGCAYTCA GCCATCYTAY TCCTGGGGAA AATGAACTT GTGCTCCTAT CAAATGCTCA 180
 5 GTTGTA AAC TGGAAAAA TTTTAGAAGA CATCTGTGCC AGCATCTGTG TTTATGTCTA 240
 TAAATGTAG AAACTAAAG CACAGAGATG TTAAATGTTT TGTCCAAGGT CCAACAGCTG 300
 10 GTTAGCARGC TTGGTCTGGT GACCTTTCTA CTGAACCACA GTGCCGCTGG GGAAGTCCT 360
 CAGCACAGAT GGCTGCTGCT ATAGCTGGGG TATGGGCAGT ATTAGTAGTT AACCAAGTCAA 420
 CCCAAGTTC CATAGTCTAG GTTCTGCTTC AGCTGGAGGT TAGGGAAAAA CACAAGAAAA 480
 15 TCCCTTACCA CTCTACCAGT GCTGGGGGAT GFACTAAGAG ATCCCC 526

20 (2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 426 base pairs

(B) TYPE: nucleic acid

25 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

30 GGCACAGGC AGGAGAGACT TGGTCCATGG GGAGAAGCCT GCAGTATAGA TGGGACCTCC 60
 AGGAGCCCAA GTAGCATAGA CCCTGCTGAT CCGGGGCCAT TGAGCCAGAG GATTGGGGCT 120
 35 GAATGTCCCC AGAGACAAA GGGAAAGTA GATCCTTTCC CTTAAAGATG AAAGCCATCG 180
 CCCGGGCTTG CTTATGCTC TCTCTCCTGG TCCTCCACA TGTGTCTTCT GAACATTGT 240
 TCTGGCATCA CAATCCCCGT CATCCTGTCA TCTGGCCCTT CCCACCTTTC CACCTTATCT 300
 40 CTTGCAGTGT CTCGGGTG ACCTGGCACC TGGGTGAARG CTTGCTCTTG CTGGTGCCCA 360
 TAGCCCCCAG TGTATGTCT TGAMCTCCCC AGCCATATGG ARACCCACCT CAGGAGGGCC 420
 45 CCTCGA 426

50 (2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 844 base pairs

(B) TYPE: nucleic acid

55 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

60 GGCACAGCG CACGAGATAG GAAGCTTGGC AGGGGCAGCT CCCCCAGTGC GCATTGCCCT 60

347

5 GTAACTCGAG CGCCTGGGAG TGGGGAGAGG CTTGGAAATG GAGCAGGGTG GTGGACCTCG 120
 TCTTCTCCTG CTCATCCCAG GCCTCCTCCA TAACACCTAC CTAGCACGGC CTGGGGACTT 180
 10 CCCAGCCCAA GGAACAACTG AGAATACTGA GTGCCAGGAT AGCCCTAGCC CCATTTTACA 240
 CCTGGGCAAA GTGAGGTGAC TGGATTCAAA CACTCAGATT TAAACCTCCT CTGTGTCTGC 300
 AGCACCTGTA TATAACTGCC AGCCTCTGCT GCCCCTCTCC AAAAAGTCTC TGCCCTTGTC 360
 15 TTTGGCACCT GTCTCTGTCC TCCCCATTCT CTGCTCCTCC TTTCTCCAAC TCAGANTCAC 420
 CCTGTTAGTT CAGCAAATGT TCATCGAGCT CCATAATGTA GCAGGACAGG NCTGTCTAAC 480
 AGATTCTGGN CTTGCAAGG TGAGACAAGT ACTCTCCATC TTTCTCTCAT CTTACAGAT 540
 GGTCTGCTCA ACAACTTTGC ACTGAATTGT AAATAATTGA TACTGCATAA AACATTGATG 600
 20 TTCTTTAAGG GTAGTCCAGC AAGGTGGCAA GTCTTATAAT GATAACTGCT CAAGGATCTC 660
 TCAGTGAAGC ATTTGGGGST GCTAGCTCTG CCTATGGGTG AGGTCAGCTA TCTCACCCA 720
 TCTACTTCCA CNTGCCCCC CATGCCAGGC TCACCTGAG CTGAGATGCC TGAGCAGGTG 780
 25 GCAGAAAGGA GCCACCTGGT TTATGCTTCG GGACCACAAA CTCCTCTATC CAGANGACAG 840
 TTTT 844

30

(2) INFORMATION FOR SEQ ID NO: 97:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1985 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

AGCCCTGCTG AAGTACAGGT TCTTCTATCA GTTCTGTGTG GGCAATGAAC GAGCAACAGC 60
 AAAGGAGATC AGGGATGAAT ATGTGGAGAC GCTGAGCAAG ATTTACCTGT CTTACTACCG 120
 45 CTCTTACCTG GGGCGGCTCA TGAAGGTGCA GTATGAGGAA GTCGCTGAGA AAGATGATCT 180
 AATGGGTGTG GAAGATACAG CAAAGAAAGG ATTCTYCTCA AAGCCATCGC TCCGCAGCAG 240
 50 GAACACCATT TTCACCTAG GAACCCGCGG CTCTGTCTATC TCCCCACTG AACTTGAGGC 300
 CCCCATCCTG GTGCCTCACA CAGCGCAGCG GNAGAGCAGA GGTATCCATT TGAGGCCCTC 360
 TTCGCAGCC AGCACTACGS CCTCCTAGAC AATCCTGCC GCGAATACCT TTTTCTGTGT 420
 55 GAATTTTGTG TTGTGTCTGG CCCAGYTGCA CACGACCTGT TCCATGCTGT CATGGGCCGT 480
 ACACTCAGCA TGACCCTGAA ACACCTGGAT TCTTATCTAG CTGACTGCTA CGATGCCATT 540
 60 GCTGTTTTTC TCTGTATCCA CATGTCTCTC CGGTTCGGTA ACATTGCAGC AAAGAGGGAT 600

	GTTCCTGCCC TGGACAGGTA CTGGGGAACA GGTGCTTGCC TTGCTATGGC CACGGTTTGA	660
5	ACTGATCCTG GAGATGAATG TTCAGAGCGT CCGAAGCACT GACCCCCAGC GCCTAGGGGG	720
	GTTCGGATACT CGGCCCCACT ATATCACACG CCGCTATGCA GAGTTCTCCT CCGCTCTTGT	780
	CAGTATCAAC CAGACAATTC CTAATGAACG GACCATGCAA TTGCTGGGAC AGCTGCAGGT	840
10	GGAGGTGGAG AATTTTGTCC TCCGAGTGGC AGCTGAGTTC TCCTCAAGGA AGGAGCAGCT	900
	TGTGTTTCTG ATCAACAACAT ATGACATGAT GCTGGGTGTG CTGATGGAGC GGGCTGCAGA	960
15	TGACAGCAAA GAGGTTGAGA GCTTCCAGCA GCTGCTCAAT GCTCGGACAC AGGAATTCAT	1020
	TGAAGAGTTG CTGTCTCCCC CTTTGGGGG TTTAGTGGCA TTTGTGAAGG AGGCTGAGGC	1080
	TTTGATGAG CGTGGACAGG CTGAGCGACT TCGAGGGGAA GAAGCCCGG TAACTCAGCT	1140
20	GATCCGTGGC TTTGGTAGTT CCTGGAAATC ATCAGTGGAA TCTCTGAGTC AGGATGTAAT	1200
	GCGGAGTTTC ACCAACTTCA GAAATGGCAC CAGTATCATT CAGGGAGCGC TGACCCAGCT	1260
25	GATCCAGCTC TATCATCGCT TCCACCGGT GCTGTCCAG CCGCAGCTCC GAGCCCTCCC	1320
	TGCCCCGGCT GAGCTCATCA ACATTCACCA CCTTATGGTG GAGCTCAAGA AGCATAAGCC	1380
	CAACTTCTGA TGTGCCAGAA ACCGCCCTGA GATCTGCCGG TCATCTCCAT GGACTTCTGC	1440
30	ACCCCATTC ATACCCCTCT TCACTGGGG TACCCCTCC AGTTTTCCTCC TTGCTTCCCA	1500
	GGCCCTTGAC ATGGCTTACC TGCCTTCACT OCCAGCACT TGCCCAACAG GATAAGCTGG	1560
35	ATCCCTTGG CCTTCTGAAT ATCCAGTGT CTTGAGGTTT CCCAAGACCA CTTCCCTGTG	1620
	GGCTTCCAAA ATGGCTTTA TCATTCTCC AGTCTGTAC CCTCCTTTC TGCTCCATA	1680
	CACCCAAGGC TTGTTCTTC CCTGTAAAA ACCACTGCCT CAATCTCTGG TTTACTCAAC	1740
40	TAGTACCAT GTCTGAGGC ATGAAGCCTC CTCAGCTCTT GGAATGCTG GCAAGGGGTG	1800
	ACTGCCTCTG AGTCATGTG TTTTCAAAG TGATTCTTT TCTGTAGCTT TTTGACCTAA	1860
45	GATCTCAGCA ATTTGAACAC TAACCTCTCC CCTCTGGCT CAAGAATTAC TCCGAAGTCA	1920
	GTCTGCAGAA AATAAATATT TAGTATGACA TGAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1980
	AAAAA	1985
50		

(2) INFORMATION FOR SEQ ID NO: 98:

- 55 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1416 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- 60

349

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

	ATATGAAGGG AAAGAATTTG ATTATGTTTT CTCAATTGAT GTCAATGAAG GTGGACCATC	60
5	ATATAAATTG CCATATAATA CCAGTGATGA CCCTTGGTTA ACTGCATACA ACTTCTTACA	120
	GAAGAATGAT TTGAATCCTA TGTTCCTGGA TCAAGTAGCT AAATTATTA TTGATAACAC	180
10	AAAAGGTCAA ATGTGGGAC TTGGGAATCC CAGCTTTTCA GATCCATTTA CAGGTGGTGG	240
	TGGTATGTT CCGGGCTCTT CGGGATCTTC TAACACACTA CCCACAGCAG ATCCTTTTAC	300
	AGGTGCTGGT CGTTATGTAC CAGGTCTGTC AAGTATGGGA ACTACCATGG CCGGAGTTGA	360
15	TCCATTTACA GGAATAGTG CCTACCGATC AGCTGCATCT AAAACAATGA ATATTTATTT	420
	CCCTAAAAAA GAGGCTGTCA CATTTGACCA AGCAAAACCT ACACAAATAT TAGGTAAACT	480
20	GAAGGAACCTT AATGGAAC TG CACCTGAAGA GAAGAAGTTA ACTGAGGATG ACTTGATACT	540
	TCTTGAGAAG ATACTGTCTC TAATATGTAA TAGTTCTTCA GAAAAACCCA CAGTCCAGCA	600
	ACTTCAGATT TTGTGGAAAG CTATTAACTG TCCTGAAGAT ATTGTCTTTC CTGCACTTGA	660
25	CATTCTTCGG TTGTCAATTA AACACCCAG TGTAATGAG AACTTCTGCA ATGAAAAGGA	720
	AGGGGCTCAG TTCAGCAGTC ATCTTATCAA TCTTCTGAAC CCTAAAGGAA AGCCAGCAA	780
30	CCAGCTGCTT GCTCTCAGGA CTTTTGCAA TTGTTTTGTT GCCCAGGCAG GACAAAAACT	840
	CATGATGTCC CAGAGGGAAT CACTGATGTC CCATGCAATA GAACTGAAAT CAGGGAGCAA	900
	TAAGAACATT CACATTGCTC TGGCTACATT GGCCCTGAAC TATTCTGTTT GTTTTCATAA	960
35	AGACCATAAC ATTGAAGGGA AAGCCCAATG TTTGTCACTA ATTAGCACAA TCTTGGAAAGT	1020
	AGTACAAGAC CTAGAAGCCA CTTTCTAGACT TCTTGTGGCT CTTGGAACAC TTATCAGTGA	1080
40	TGATTCAAAT GCTGTACAAT TAGCCAAGTC TTTAGGTGTT GATTCTCAAA TAAAAAGTA	1140
	TTCTCAGTA TCAGAACCAG CTAAAGTAAG TGAATGCTGT AGATTATCC TAAATTGCT	1200
	GTAGCAGTGG GGAAGAGGGA CGGATATTTT TAATTGATTA GTGTTTTTTT CCTCACATTT	1260
45	GACATGACTG ATAACAGATA ATTAAAAAAA GAGAATACGG TGGATTAAAGT AAAATTTTAC	1320
	ATCTTGTAAG GTGGTGGGGA GGGGAAACAG AAATAAAATT TTTGCACTGC TGAAAAAAA	1380
50	AAAAAAAAA AAAAGGAAAC TCGAGGGGGG GCCCGG	1416

(2) INFORMATION FOR SEQ ID NO: 99:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1935 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

5	NICTACCTA ATCAAGATGG GGACATACTT CGGACCAGG TTCTTCATGA ACATATCCAG	60
	AGATTGTCTA AAGTAGTGAC TGCAAATCAC AGAGCTCTTC AGATACCAGA GGTTTATCTT	120
	CGAGAAGCAC CATGGCCATC TGCACAATCA GAAATCAGGA CAATAAGTGC TTATAAAACC	180
10	CCCCGGGACA AAGTGCAGTG CATCCTGAGA ATGTGCTCTA CGATTATGAA CCTCCTGAGC	240
	CTGGCCAATG AGGACTCTGT CCCTGGAGCG GATGACTTTG TTCCTGTGTT GGTGTTTGTG	300
	TTGATAAAGG CAAATCCACC CTGTTTGCTG TCTACTGTGC AGTATATCAG TAGCTTTTAT	360
15	GCTAGCTGTC TGTCTGGAGA GGAGTCCTAT TGGTGGATGC AGTTCACAGC AGCAGTAGAA	420
	TTTATTAAAA CCATCGATGA CCGAAAGTGA CCAAGACCAA GGCCACCAA GGCAGCAGAC	480
20	TGTTAATCAG ACAACAGAT CTCTGAGAAG GTGCATCAGC TGCTTTGAAG GCTGAAGATT	540
	GTTTTGTATG ATACTGCACA GCATCAGGCA TTTTAAAGCA GATCTTTACT AAACAGGTTA	600
	ATGAGCTAAC AAGCAGGTC TCTCGTCTTT GGGCTCTTTC CTTCTGAGT TGCATATTCT	660
25	ATTTTCTTGT CCCCAAGTAG AGACTAGTAC TACAAAAAGG GACCACATTT TTCAAGTATT	720
	TCTAAGTATA AAAACAAAA CAAAATCTC TTAGGAAATG TCTAGACCTC CATTCTTGGA	780
30	TTCCCTTTCT TTCTTTTAT TTTAAAAAG AACAGTACCC CTCTTTTAAG ATGCTGTCTT	840
	ACATTAATGA GCATCTAATG GAAAGAAGGT ATGAGTTGCA CTGAGGATTA GAATAGTGGT	900
	GGGTTAGTGG CATTATCTAT AAATACACTC ACCTAAATTG AAAGCTAAGA AGGAAATGTA	960
35	AATATAATAT ATATTTATAT TTGATGTAAT ATGGACATCT GCAGATTCTA ATAAACAAGG	1020
	ACTATTGCTG ATAGTAGGCT GTGACATACT GTCTTGTAAG ATGGTTTCCT TGACAAAATT	1080
40	TAAGCTGAGC TTAAAGCAA AAAACAAAA AGTACACAGA AATATTTATT AAAATGTAAT	1140
	ACAGTTTATT GAACCTTCTA GGTATGGAGT TTGATGGACA GGGCTGCCTY TAATGAGTGT	1200
	GAAGGTCACT AAGTCACTTA GACATCTCAC CGTGAAGTT TGTGAGCCTG CATTAGGAGA	1260
45	TAGACTGATT ACCATACATG ACATAAAAAG GAACAGTGA TAGCTCATAC TTTATGGTGG	1320
	TTCTTCTCCT CCGAAATAAT ATACTGCAGA AATCCCAGAC AGAGCTCCTT ACAAACCTTT	1380
50	AATTGTAATA TATTTTGTAT GATTATTCAC ATTGAATGCA CAGACCAAGA ATTCAGTGAA	1440
	TGTCATTTTT TAAAAACTA ATTTGTATTG TCTGCTCTAG TGATACAAGT TTTACTAGTG	1500
	ATAAACTATT TTAATCAACC ATACTATTCT TATGAAAAA AATATCTATT TTGGCAGGTT	1560
55	TCTGTGCCTT TATTTCCCTC TTCTGAAAAA AAGTCTGTGT TTTTCATAGTT TGGTTTGCAT	1620
	TGTATATCAA TAATTAATCA GGAATGGGTT TTGGTGCTG AAAAATTGGC CATGGAGGCA	1680
60	CACCAAAGCT TCAAGCACA GTCTGTGACA TGGGCCATCA CTGCTGGTT TCACTTCGTG	1740

5 TGTTTCCTAA ACACATTTAG CTGCTTTTTT AACAACTCA GCCCCATACT TGAGTCCCTT 1800
 GTTGTGGGA GCATTCCAG GCATCTTTA AGGGAAGTGT GACAAACAGC CTCGGGCAGA 1860
 TGAACACGGA GGCTCTCTGT TGTCTGTCTC TGAGATCTTT GTGTCTGGGA ATGCCTAAAG 1920
 NTTTTGNTTT TTTTT 1935

10

(2) INFORMATION FOR SEQ ID NO: 100:

15

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 599 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

GAATTCGGCA CGAGCGTCCA CGCAGCCGCC GGCGGCCAG CACCCAGGGC CCTGCATGCC 60
 25 AGGTGCTGG AGGTGCAGC GAGACATGCA CCCGCCCGG AAGCTCCTCA GCCTCCTCTT 120
 CCTCATCTG ATGGGCACTG AACTCACTCA AGACTCCGCT GCGCCCGACT CCCTGCTGAG 180
 AAGTCAAAG GGCAGCAGCA GGGGGTCTTT GGCTGCTATT GTCATCTGGA GGGGAAGAG 240
 30 TGAGAGCCGG ATAGCCAAGA CCCAGGCAT TTTCAGAGGT GCGGGACCT TAGTCCTACC 300
 CCCAACACAC ACCCCTGAGT GGCTCATCCT CCCTTTGGGC ATAACGCTGC CCTTGGGGGC 360
 35 TCCAGAAACA GCGGTGGGG ATTGTGCCGC TGAGACCTGG AAGGGCAGCC AGCGTGCCGG 420
 CCAGCTGTGT GCATTGCTGG CTTAATATGC AGGGCTTGGG GGGCTGTGGC CACATGCCCG 480
 GCAGGAGTGT AGTGAGGAGC CCTGTGGCGT GCTGGTGTGG GGATCGTGGG CATTTCAAAC 540
 40 GGGCTGTGCG TACCCTGAAC AATGTATCAA TAGAGAAAAA AAAAAAAAAA AAAACTCGA 599

45

(2) INFORMATION FOR SEQ ID NO: 101:

50

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 784 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

GAATTCGGCA CAGAAAAAA AGAGAGACTG GGTCTTACTG TGTGCCCAG ACTTGTCTTG 60
 AACTCCTGCC TCAGCCTCTC AAGTACTTGG GATTATAGGC CAAGAAGCCA CCATGCCTAG 120
 60 CTTCTTCTG TCATTGATCC AGACTAATAC TCTGGGTCA GCCTCATPTC TTCTCTTTCT 180

CACTTTGCAC ATCCACTTGT CACCAAATCK RGTTCATTCT GCATCCTAAG TAAGTCCTTT 240
 GATTCTCTCA GTTGTTTCATT AGTAATGTCT CAARTGTAAT TTTTCTAGT AGTTTTCAGC 300
 5 CTGTCTTTCC KGCCTTCAGT CTTAACTTCT CCAGTACATA KGCCACATG TTGTCAGCAK 360
 GATCAWATTT TATTTAAAAA TACTTTACAW AKGTTTATKG CCAAATATTA GRAAATACAG 420
 10 ATTCATGGAA AGAAAAATCA CTGTCCCAAG GAGGTCACCTG GCATGGTGAG GTTAAGGGGT 480
 GATTTTAATT TTAAAAATG TATATTTTTT CCTGTGTAGA GTAGTAACAC CCTTGAAAAC 540
 ACAWTCCCTT GTAAAGTCTC TAATTCTGTA CTCCGCATCT AGSTGRTCTC TTCTTTCTCA 600
 15 GATATTTTAC AATTTTCATT ATCACCACCT TTCTCTAGCC TTACCCGTC TCTTCAATAT 660
 TWACATATGC AGAAGTTTCT CCTAACAAAC ACCTGCCTCT GCCTCAGTTC TGCTACCACC 720
 20 CTGTGTCTTT CTTTCCCTTC ACAATCAAAT TTAAGAGTGT CAAAAAAAAA AAAAAAAAAAC 780
 TCGA 784

25

(2) INFORMATION FOR SEQ ID NO: 102:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1035 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

AGAGGCTTGG CTGCGTGGCC CTATCTCCGT CTCGCCACC CACTTAGCGT TTAGGCATC 60
 AATTACCAGC AGTTTCTCCG CCACTATCTG GAAAAITACC CGATTGCTCC CGGCAGAATA 120
 40 CAAGAGCTTG AAGAAGCCCG CAGTTGCGTG GAAGCCTGCA GAGCAAGGGA AGCAGCGTTT 180
 GATGCCGAAT ATCAGCGAAA TCCTCACAGG GTGGACCTCG ATATTTTAAC CTTTACGATA 240
 45 GCTCTGACTG CCTCTGAAGT TATCAACCTT CTGATAGAAG AACTTGGTTG CGATAAGTTT 300
 ATCAATAGAG AATAGTTAGG TGGTGACACT ACTTCAAGAG AACCTCTGCA TTCCAGTCAT 360
 ACCAATCCTG CAACTTGATT TTCAGAAGTC AAGAGTATAT CGCGATAAGA CAGTGCACAG 420
 50 GTGGAGGGGA AAAAAAGGGG GAGGGGAAG CTTATCTTGA AAAAGCATCA CAGAAGTAGA 480
 AAAAAATGTC GAAAGCATTA TAACTGTAAC GTTCTTTGAG TTGTGTATTG ATCCACATTT 540
 55 TTCCCCCTGC ATTATGGAAA ATGTCTCTCA GCATTGCTTT ATTACAAAGT AAAGGATGGT 600
 TTTATAAAAT TGAGACTGAT GAAACATCAA TACTAGAGCC CATGAGGATG AAAGAAATTA 660
 TCAAATAGTG CTGAACAGAA TAAGATGTTA ACGCTGAGTT ATTAGGACTG GAAGGCTATG 720
 60

353

AAAAGAACTT GAAATTGTGC GAATATGTGC TCTCTTCATG TCATATTCAA TAGAAGTTTC 780
 TAGTTTAAGA TTGATTTTGT GTTTTCTTAG GCATTTCAAG TGACAAGCAA AGTAAATGTA 840
 5 TATATTATGT GATAAATCAT GTTTTCAAGA ACGTCAAATT TCTGGACTTT TTTCTTTCAA 900
 TTTTAAATTT TTAAAGTTT TTTGGTATTA AAAAATCYAT TCACAAGCCA AAAAATWTWT 960
 WAAATWTWCM GCGAAAAGCC AAAAAAAAAA AAAAMMAGGG GGGGCCGGGC CCCATCCCCC 1020
 10 CAAGGGGGTC CNGNT 1035

15

(2) INFORMATION FOR SEQ ID NO: 103:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2218 base pairs
 20 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

25 AGGTATTAGG CCTTTTGTG GGAGCCCCAT GTTTGTGTTT TCTGAGTTGG TGGGGAGGGA 60
 SGGAGGGGGA GGGCTGAATT GTTTTCAGA GGAAGATGCG ATCTGTGCTT TAAATTTCTC 120
 30 ATTACTGGGT TAGAAAACAA AGAGGGAKTG CCCTGCACAT TTTCTTTTGT GCTTTTAAAT 180
 GTTCTTAAG TTGGAACAGG TTTCCTCGGG CCTGTTTGA CTGATTGCTG GAGTGCATTT 240
 GATAGTTAAA AATTACTAAT TGGTTTATT TCCCTTCACA CTCTGCCTCC CCACTTCTCC 300
 35 CCCCCTTACT GAAAATAAC CATTTTAGTG TCAGGCTAGA AATTGAATG CTGAGTTTTC 360
 TGTATCCTTT AAATTAAAA CCACAAGTGT TTATTGTAGT GGTAAACTG TAGCATCTCA 420
 40 GCATCTGGGT GGAAGCTGCC TATATTCTTT CCCAGTTTAA CTGGGGACCA TCTGTGAAAT 480
 TAATTTTCCA TCCAGACAGC TGCTGTGAGC AAATGAACAT AAATGCTCGC TGGAAATTTA 540
 CTAACCAATT TTTATATTGA CCTGCAGTGT AAAAAGCACA TTTAATTATA AACAATATAT 600
 45 TCAAAATGGG CAAATTTTAT TTTCAAATGC AGTGTAGAGC TAGATTAAAA GCAACTCTTT 660
 GCCACCTACT CTGCCCTTTT GGCAAAGTTA CCTTGAACAA AGAATCTTAA GGGTTTATTA 720
 50 AGAACTCTTT ATTTCTTCA TACCCGTGTC TCTGCAGTGC TTTCTAACAG CTCTGGGTG 780
 CAGATTTTCT TCGCATCCT TTTGCACTCA GCTTATTACA GGTAGGTAGT GCTTAAGAAA 840
 AGTCATGGAG GACTAAAGCC TAAGTCCTTT TCACTTTTCC TCCATCTGAA GGTAGGTGAG 900
 55 TTCATCTCT TCATAGTAAT GCTGTTTAC CAAGACTTTA TAGCAGATGG ACCCAGAAAG 960
 AATTTTCTGC TATTGTGTT ACTACAACAG GATAGGGACA TCAGACAGCC CCAGAAACCC 1020
 60 CTCCAGATC TGATATGGGA CTATTAATTT TTATGCTGTT AATTGGTATT CATTCACAAT 1080

	GCAGTTGAAG GGGGAAGGCT CCACTGCAIT CTTTGGCTAA GGCCTGAATG CTTGCTCATC	1140
5	TGTAAGATCT ATACTCGAGG TTTTGTTC CTTTAAAAT TCTTTAGGGA GAGAGGGATG	1200
	GTTTCTGAGG GGTTCGAAA GTATGATTCA ATGTGCAACA TACAGGTAGG TCTTCAGCAT	1260
	AAGCTGAAAT ATATGCATGT AAAAAGTTG ACATCTTTTT TTTTAATTTT CCACTTTCIT	1320
10	CTTAACITTA CTCTCTTTT TGTCCTCCCC CCATCTTACA GAAGTTGAGG CCAAGGGAGA	1380
	ATGGTAGGCA CAGAAGAAAC ATGGCAAAC GCTCTGTGCT TTCAAACCAA AGTGTTCCTC	1440
15	CCAACCCCAA ATTTGTCTAA GCACTGGCCA GTCTGTTGTG GGCATTGTTT TCTACAACCA	1500
	AATTCGGGT TTTTTCCTC TTTCTTAAA CATAGAGGTA CCACCACAAG GGATGCCCTA	1560
	CTCTCTCGCA GCTCTGAAA GCATCTGTTT GAGGAAAGG TCTCTGGCA AGCAAGTGGT	1620
20	TATTTGGATT GCTTGCTTCC CTTTTCAC CTGGGACAT GYAATCATAA AATAACAGTA	1680
	AATCCAAAC CTCAAAACT ATTATGGCT GAGCACAGCT GAAATCTAGC AGAGTTTAAC	1740
25	TCTCTGCCT CCATGCTGT CACTTATAAT TCAGTTCTG CTGTTGGCTT CAGAACATGA	1800
	GCAGAAGAT CGTTTATGC TAGTTATGC ATTCATGGTT GAAACTCAAC TTAGGGAAAG	1860
	GGTCCAATG TATTAAGCAA TGGGCTGCTT CTCCCCAATC CTCCCTAACA ATTCTGTG	1920
30	TGGACTTCTC ATCTAAAAGG TTAGTGGCTT TTGCTGGGA TCAGTCTCT CTATTGATGT	1980
	TCTTGCTGGT CTCCAGACAC ATTCCTGTG CATTAAGACT TGAAAGACTT GTAGATGTGT	2040
35	GATGTTTCTG CACAGGATGC TGAAAGCTAT GTTACTATC TTAGTTGTA AATTGTCCTT	2100
	TTGATACCAT CATCTGTTT TCTTTTGTA GGTATAATA AAAACACTGT TGACAATAAA	2160
40	AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	2218

(2) INFORMATION FOR SEQ ID NO: 104:

- 45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1351 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

- 50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

	CTTCACAGAC TGACAGAATG GTTTTGTTC GTTTTGTTC GTTTTGTTC GTTTTGTGAGA	60
55	TGGACTCTAG CTCTGTCAAC CAGGCTGGAG TGCAGTGGT CGATCTCGGC TCACTGCAAG	120
	CTCCGCTCC CGGTTCTCA CCATCTCTCT GCCTCAGCCT CCCGAGTAGC TGGGACTACA	180
60	GGCGCCCAAC ACCACGCCG GCTAATTTTT TGTATTTTTT AGTAGAGACG GGGTTTCACC	240

355

	ATGTTAGCCA GGATGGTCTC GATCTCCTGA CCTCGTGATC CGCCCGCYTC GGCCTCCCAA	300
	AGTGCTGGGA TTACAGGCGT GAGCCACCGT GCCTGCCCA GAATGGTTT TAAAGCCACA	360
5	GTTGAGARGC CACCCATTGC CCGCGCCTG GACAGTGATC ATCTTGTTCA TCTTGTTTCA	420
	TCCTTTCTTG TGTGATTGGA ATTATTATC CCCTTTGAAA GATGAGAAGG TTGAGATGCA	480
10	AAGAGTCTAC CTTTCCAAGT TCTCACTGCT GGAAAGARCT AGAAGCACAG TTCAAAGTTC	540
	TGGMITCTGG ACTCTGCAGT CCAGGTYTCC CTTYTCCAC TTGCCTACCC TCAATGCCAC	600
	ACTGTTTTTG AAGTGGCCCA TAACTGAAG GRAAAGTTTA AAGACAGTTC AATTTAATCA	660
15	TCAGRATGCA TTCTTTTTT TTTGGARAC GGARTTTCAC TCTTGCTGCC CAGCTGGAG	720
	TGCAATGGTG CAATGATCTC GGCTCACTGC AACCTATGCC TCCTGGGTTT AAGNGATTAT	780
20	CCAGCCTCAG CCTCCCGAGT AGCTGGGATT ATGGGCGCCC ACCACCATGC CCAGCTAATT	840
	TTTGATTTTT TTTTTTAGT AGAGATGGG TTTGCCAGG TTGGCCAGG TGTCTTGTG	900
	AAATCCTGGC YTCAGTGAT YTGCCACYT CATCTCCAA AAGTGCTGGG ATTACAGGCA	960
25	TGAGCCACTG CGCCTGGCYT CAGAATGCAT TCTTACACAT CTATCCTAGA CATTTATAAG	1020
	CACTCTAATG GATAACAATC CAAGAATAAA TGATTGTAAA AGATGATGCC GAAGAGTTGA	1080
30	TGTCATCTT TTTTCTCTAA GAAAAAAGT CCGGAGTAT TAAATATTTA GATCAATGTT	1140
	TATAAATGA TTAATTTGTA TATCTCATTA TTCTATTTT GGAATAAAAA CTGACCTTCT	1200
	TTAATCATAT ACTTGCTTT TGTAAATAGC AGCTTTGTG TCATCTCTCC CACTTTATTA	1260
35	GTTAATTTAA ATTGGAAAA ACCCTCAAAC TAATATCTT GTCTGTTCCA GTCTTATAAA	1320
	TAAACTTAT AATGCATGTA AAAAAAAAAA A	1351

40

(2) INFORMATION FOR SEQ ID NO: 105:

- 45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2066 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

- 50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

	GGCAGGAGGC GCGGAGGGC CACAATCACA GCTCCGGGCA TTGGGGGAAC CCGAGCCGGC	60
55	TGCGCCGGGG GAATCCGTGC GGGCGCCTTC CGTCCGGTTC CCATCCTGCG CGCGCTCCAG	120
	CACCTCTGAA GTTTTGACGC GCCCAGAAAG GAGGCGAGGA AGGAGGGAGT GTGTGAGAGG	180
	AGGGAGCAAA AAGCTCACCC TAAACATTT ATTTCAAGGA GAAAAGAAAA AGGGGGGGCG	240
60	CAAAAATGGC TGGGGCAATT ATAGAAAACA TGAGACCAAA GAAGCTGTGC ATTGTTGGTG	300

	GGATTCTGCT CGTGTTCCTAA ATCATCGCCT TTCTGGTGGG AGGCTTGATT GCTCCAGGGC	360
	CCACAACGGC AGTGTCCTAC ATGTCGGTGA AATGTGTGGA TGCCCGTAAG AACCATCACA	420
5	AGACAAAATG GTTCGTGCCT TGGGGACCCA ATCATTGTGA CAAGATCCGA GACATTGAAG	480
	AGGCAATTCC AAGGGAAATT GAAGCCAATG ACATCGTGTT TTCTGTTCAC ATTCCCCTCC	540
10	CCCACATGGA GATGAGTCCT TGGTTCCAAT TCATGCTGTT TATCCTGCAG CTGGACATTG	600
	CCTTCAAGCT AAACAACCAA ATCAGAGAAA ATGCAGAAGT CTCCATGGAC GTTTCCTTGG	660
	CTTACCGTGA TGACGCATTT GCTGAGTGGG CTGAAATGGC CCATGAAAGA GTACCACGGA	720
15	AACTCAAATG CACCTTCACA TCTCCCAAGA CTCCAGAGCA TGAGGGCCGT TACTATGAAT	780
	GTGATGTCTT TCCTTTCATG GAAATGGGT CTGTGGCCCA TAAGTTTAC CTTTTAAACA	840
20	TCCGGCTGCC TGTGAATGAG AAGAAGAAAA TCAATGTGGG AATTGGGGAG ATAAAGGATA	900
	TCCGGTTGGT GGGGATCCAC CAAAATGGAG GCTTCACCAA GGTGTGGTTT GCCATGAAGA	960
	CCTTCCTTAC GCCCAGCATC TTCATCATTG TGGTGTGGTA TTGGAGGAGG ATCACCATGA	1020
25	TGTCCCGACC CCCAGTGCCT CTGGA AAAAG TCATCTTTGC CCTTGGGATT TCCATGACCT	1080
	TTATCAATAT CCCAGTGGAA TGGTTTCCA TCGGGTTGA CTGGACCTGG ATGCTGCTGT	1140
30	TTGGTGACAT CCGACAGGGC ATCTTCTATG CGATGCTTCT GTCCCTCTGG ATCATCTTCT	1200
	GTGGCGAGCA CATGATGGAT CAGCAGGAGC GGAACCATAT TGCAGGGTAT TGGAAGCAAG	1260
	TCGGACCCAT TGCCGTGGC TCCTTCTGCC TCTTCATATT TGACATGTGT GAGAGAGGGG	1320
35	TACAACTCAC GAATCCCTTC TACAGTATCT GGAATACAGA CATTGGAACA GAGCTGGCCA	1380
	TGGCCTTCAT CATCGTGGCT GGAATCTGCC TCTGCCTCTA CTTCTGTGTT CTATGCTTCA	1440
40	TGGTATTTCG GGTGTTTCGG AACATCAGTG GGAAGCAGTC CAGCCTGCCA GCTATGAGCA	1500
	AAGTCCGGCG GCTACACTAT GAGGGGCTAA TTTTTAGGTT CAAGTTCCTC ATGCTTATCA	1560
	CCTTGGCCTG CGCTGCCATG ACTGTCTCTT TCTTCATCGT TAGTCAGGTA ACGGAAGGCC	1620
45	ATTGGAATG GGGCGGCGT ACAGTCCAAG TGAACAGTGC CTTTTTCACA GGCATCTATG	1680
	GGATGTGGAA TCTGTATGTC TTTGCTCTGA TGTTCCTGTA TGCACCATCC CATAAAACT	1740
50	ATGGAGAAGA CCAGTCCAAT GGAATGCAAC TCCCATGTAA ATCGAGGGAA GATTGTGCTT	1800
	TGTTTGTTC GGAACCTTAT CAAGAATTGT TCAGCGCTTC GAAATATTC TTCATCAATG	1860
	ACAACGCAGC TTCTGGTATT TGAGTCAACA AGGCAACACA TGTTTATCAG CTTTGCATTT	1920
55	GCAGTTGTCA CAGTCACATT GATGTACTT GTATACGCAC ACAAATACAC TCATTAGCC	1980
	TTTATCTCAA AATGTTAAAT ATAAGGAAAA AAGCGTCAAC AATAAATATT CTTGAGTATA	2040
60	AAAAAAAAA AAAAAAAAAA AAAAAA	2066

5 (2) INFORMATION FOR SEQ ID NO: 106:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1705 base pairs

10 (B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

15 AATTGGCAK AGGGCAGCTG TCGGCTGGAA GGAAGTGGTC TGCTCAGCT TGCTGGCTTG 60

CGCATCAGGA CTGGCTTTAT CTCTGACTC ACGGTGCAAA GGTGCACTCT GCGAACGTTA 120

AGTCCGTCCC CAGCGCTTGG AATCTACGG CCCCCACAGC CGGATCCCTT CAGCCTTCCA 180

20 GGTCTCTAAC TCCCGYGGAC GCTGAACAAT GGCTCCATG GGGCTACAGG TAATGGGCAT 240

CGCGCTGGCC GTCTGGGCT GGCTGGCCGT CATGCTGTGC TCGCGCTGC CCATGTGGCG 300

25 CGTGACGGCC TTCATCGGCA GCAACATTGT CAGCTCGAG ACCATCTGGG AGGGCCTATG 360

GATGAAGTGC GTGGTGCAGA GCACCGGCCA GATGCAGTGC AAGGTGTACG ACTCGCTGCT 420

GGCACTGCCG CAGGACCTGC AGGCGGCCCG CGCCTCGTC ATCATCAGCA TCATCGTGGC 480

30 TGCTCTGGGC GTGCTGTGT CCGTGGTGGG GGGCAAGTGT ACCAAGTGC TGGAGGATGA 540

AAGCGCCAAG GCCAAGACCA TGATCGTGGC GGGCGTGGTG TTCTGTGTGG CCGGCCCTTAT 600

35 GGTGATAGTG CCGGTGTCTT GGACGGCCCA CAACATCATC CAAGACTTCT ACAATCCGT 660

GGTGGCTTCC GGGCAGAAGC GGGAGATGGG TGCTCGCTC TACGTGGCT GGGCCGCTC 720

CGGCTGTCTG CTCTTGGCG GGGGGCTGCT TTGCTGCAAC TGTCCACCCC GCACAGACAA 780

40 GGCTTACTCC GCCAAGTATT CTGCTGCCCC CTCTGTGCT GCCAGCACT ACGTGTAAAG 840

TGCCACGGCT CCACTCTGTT CCTCTGTGCT TTGTTCTTCC CTGGACTGAG CTCAGCGCAG 900

45 GCTGTGACCC CAGGAGGGCC CTGCCACGGG CCACTGGCTG CTGGGGACTG GGGACTGGGC 960

AGAGACTGAG CCAGGCAGGA AGGCAGCAGC CTTGAGCTC TCTGGCCAC TCGGACAACT 1020

TCCCAAGGCC GCTCTCTGCT AGCAAGAACA GAGTCCACCC TCCTCTGGAT ATGGGGAGG 1080

50 GACGGAAGTG ACAGGGTGTG GTGGTGGAGT GGGGAGCTGG CTTCTGTGCT CCAGGATGGC 1140

TTAACCCTGA CTTTGGGATC TGCTGCATC GGTGTGGCC ACTGTCCCA TTTACATTTT 1200

55 CCCCACCTG TCTGCCTGCA TCTCTCTGT TGCGGGTAGG CCTTGATATC ACCTCTGGGA 1260

CTGTGCCTTG CTCACCGAAA CCGCGGCCA GGAGTATGGC TGAGGCCTTG CCCACCCACC 1320

TGCCTGGGAA GTGCAGAGTG GATGGACGGG TTTAGAGGGG AGGGGCGAAG GTGCTGTAAA 1380

60

358

CAGGTTTGGG CAGTGGTGGG GGAGGGGGCC AGAGAGGCGG CTCAGGTTGC CCAGCTCTGT 1440
 GGCTCAGGA CTCTCTGCCT CACCCGCTTC AGCCAGGGC CCCTGGAGAC TGATCCCCTC 1500
 5 TGAGTCTCTT GCCCTTCCA AGGACACTAA TGAGCCTGGG AGGGTGGCAG GGAGGAGGG 1560
 ACAGCTTCAC CCTTGAAGT CTTGGGGTTT TTCTCTTCC TTCTTTGTGG TTTCTGTTT 1620
 10 GTAATTTAAG AAGAGCTATT CATCACTGTA ATTATTATTA TTTTCTACAA TAAATGGGAC 1680
 CTGTGCACAG GRAAAAAAAA AAAAG 1705

15

(2) INFORMATION FOR SEQ ID NO: 107:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1167 base pairs
 20 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

TGCAGGAATT CGGCAGAGT TTTCCGCTAG ACTCTGGCAG TTGGTGAGCA TCATGGCAAC 60
 CGTTACAGCC ACAACCAAAG TCCCGGAGAT CCGTGATGTA ACAAGGATTG AGCGAATCGG 120
 30 TGCCCACTCC CACATCCGGG GACTGGGGCT GGACGATGCC TTGGAGCCTC GGCAGGCTTC 180
 GCAAGGCATG GTGGGTGAGC TGGCGGCACG GCGGGCGGCT GCGGTGGTGC TGGAGATGAT 240
 CCGGGAAGGG AAGATTGCCG GTCGGGCAGT CCTTATTGCT GGCCAGCCGG GCACGGGGAA 300
 35 GACGGCCATC GCCATGGGCA TGGCGCAGG CCTGGGCCCT GACAGCCAT TCACAGCCAT 360
 CGCCGGCAGT GAAATCTTCT CCCTGGAGAT GAGCAAGACC GAGGCGCTGA CGCAGGCCTT 420
 40 CCGGGCGTCC ATCGGCGTTC GCATCAAGGA GGAGACGGAG ATCATCGAAG GGGAGGTGGT 480
 GGAGATCCAG ATTGATCGAC CAGCAACAGG GACGGGCTCC AAGGTGGGCA AACTGACCCT 540
 CAAGACCACA GAGATGGAGA CCATCTACGA CCTGGGCACC AAGATGATTG AKTCCCTGAC 600
 45 CAAGGACAAG GTCCAGGCCG GGGACGTGAT CACCATCGAC AAGGCGACGG GCAAGATCTC 660
 CAAGCTGGGC CGCTCCTTCA CACGCGCCCG CGAACTACGA CGTATGGGC TCCAGACCA 720
 50 AGTTGCTGCA GTGCCAGAT GGGGAGCTCC AGAAACGCAA GGAGGTGGTG CACACCGTGT 780
 CCCTGCACGA GATCGACGTC ATCAACTCTC GCACCCAGGG CTTCCTGGCG CTCTTCTCAG 840
 GTGACACAGG GGAGATCAAG TCAGAAGTCC GTGAGCAGAT CAATGCCAAG GTGGCTGAGT 900
 55 GGCAGGAGCA GGGCAAGGCG GAGATCATCC CTGGAGTGCT GTTCATCGAC GAGGTCCACA 960
 TGCTGGACAT CGAGAGCTTC TCCTTCCTCA ACCGGGCCCT GGAGAGTGAC ATGGGCGCTG 1020
 60 TCCAGCAGGT CTATGGGGAT GCCGTGAGGG CTCTGGTAGC TGGTGCCCG GATTGCGGTG 1080

ATGCCACGGT TGGTGGCCTC GTGCCGAATT CCTGCAGCCC GGGGGATCCA CTAGTTCTAG 1140
AGCGGCCGCC ACCGCGGTGG ANCTCCN 1167

5

(2) INFORMATION FOR SEQ ID NO: 108:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1907 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

15

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

20

GGCACAGGGG AATCATCGTG TGATGTGTGT GCTGCCTTTG TGAGTGTGTG GAGTCCTGCT 60

CAGGTGTTAG GTACAGTGTG TTTGATCGTG GTGGCTTGAG GGAACCCCTT GTTCAGAGCT 120

GTGACTGCGG CTGCACTCAG AGAAGCTGCC CTGGCTGCT CGTAGCGCGG GGCCTTCTCT 180

25

CCTCGTCATC ATCCAGAGCA GCCAGTGTCC GGGAGGCAGA AGGTACCGGG GCAGCTACTG 240

GAGGACTGTG CGGGCCTGCC TGGGCTGCCC CTTCCGCGGT GGGGCCCTGT TGCTGCTGTC 300

30

CATCTATTTC TACTACTCCC TCCCAAATGC GGTGGGCCCG CCCTTCACTT GGATGCTTGC 360

CCTCTGGGC CTCTCGCAGG CACTGAACAT CTTCTGGGC CTCAAGGGCC TGGCCCCAGC 420

TGAGATCTCT GCAGTGTGTG AAAAAGGGAA TTTCACGTG GCCCATGGGC TGGCATGGTC 480

35

ATATTACATC GGATATCTGC GGCTGATCCT GCCAGAGCTC CAGGCCCGGA TTGAACTTA 540

CAATCAGCAT TACAACAACC TGCTACGGGG TGCACTGAGC CAGCGGCTGT ATATTCTCCT 600

40

CCCATGGAC TGTGGGGTGC CTGATAACCT GAGTATGGCT GACCCCAACA TTGCTTCTCT 660

GGATAAAGTG CCCCAGCAGA CCGGTGACCG TGCTGGCATC AAGGATCGGG TTACAGCAA 720

CAGCATCTAT GAGCTTCTGG AGAAGGGGCA GCGGGCGGGC ACCTGTGTCC TGGAGTACGC 780

45

CACCCCCTTG CAGACTTTGT TTGCCATGTC ACAATACAGT CAAGCTGGCT TTAGCGGGGA 840

GGATAGGCTT GAGCAGGCCA AACTCTTCTG CCGGACACTT GAGGACATCC TGGCAGATGC 900

50

CCCTGAGTCT CAGAACAACCT GCCGCCTCAT TGCTTACCAG GAACCTGCAG ATGACAGCAG 960

CTTCTCGCTG TCCCAGGAGG TTCTCCGGCA CTTGCGGCAG GAGGAAAAGG AAGAGGTTAC 1020

TGTGGGCAGC TTGAAGACCT CAGCGGTGCC CAGTACCTCC ACGATGTCCC AAGAGCCTGA 1080

55

GCTCTCATC AGTGGAAATGG AAAAGCCCT CCCTCTCCGC ACGGATTTCT CTTGAGACCC 1140

AGGGTCACCA GGCCAGAGCC TCCAGTGGTC TCCAAGCCTC TGGACTGGGG GCTCTCTTCA 1200

60

GTGGCTGAAT GTCCAGCAGA GCTATTCTCT TCCACAGGGG GCCTTGCAGG GAAGGGTCCA 1260

360

5 GGA CTG GAC TCT TAA GAT G CGT CTT GTG CCG TTT GGG CCA GTC ATT TTC C CTCTCTGAGC 1320
 CTCGGTGTCT TCAACCTGTG AAATGGGATC ATAATCACTG CCTTACCTCC CTCACGGTGTG 1380
 10 TTGTGAGGAC TGAGTGTGTG GAAGTTTTC ATAAACTTTG GATGCTAGTG TACTTAGGGG 1440
 GTGTGCCAGG TGTCTTTCAT GGGGCCCTTC AGACCCACTC CCCACCCTTC TCCCCTTCCT 1500
 TTGCCCCGGG ACGCCGAAC CTCTCAATGG TATCAACAGG CTCCTTCGCC CTCTGGCTCC 1560
 TGGTCATGTT CCATTATTGG GGAGCCCCAG CAGAAGAATG GAGAGGAGGA GGAGGCTGAG 1620
 TTTGGGGTAT TGAATCCCC GGCTCCACC CTGCAGCATC AAGGTTGCTA TGGACTCTCC 1680
 15 TGCCGGGCAA CTCTTGCGTA ATCATGACTA TCTCTAGGAT TCTGGCACCA CTTCTTCCC 1740
 TGGCCCCTTA AGCCTAGCTG TGTATCGCA CCCCCACCC ACTAGAGTAC TCCCTCTCAC 1800
 TTGCGGTTTC CTTATACTCC ACCCCTTCT CAACGGTCCT TTTTAAAGC ACATCTCAGA 1860
 20 TTAAAAAAA AAAAAAAAA AAAAAAAAAA AAAAAAGGG CGGCCGC 1907

25

(2) INFORMATION FOR SEQ ID NO: 109:

30

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 611 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

40

45

50

55

ATGAATTAAC GCCAAGCTNT NAATAGGGAC TCACTATGGG GGAAAGNTGG GTAACGCCCTG 60
 CAGGTACCGT TCCGGAATTC CCGGGTCGAC CCACGCGTCC GATGGGGCTT TAGTAAATCA 120
 GGCTTGCAGG CTCAAAGCTG CAATCTGCCC ACTCTCAGGT ACTGAGACTT TGTGGGCCTC 180
 AGACACCAGG AAGAAAGTTG GGATACAGTC ATTTGAGTTA AAAAGGGAAT GACCCCTCAG 240
 AAACCCGCAT TAGCAGTGTT ACTCTTGGA GTGCCCTTAC TTTTAAOGCT CTCTGTTCTG 300
 AAAAAGAGGT GTTTGGTTAC GTGTGAGCCA ACATCACGTT TGTGTAGCTG TGATTACCT 360
 TTGTCGGTTT AAAAGACTTC ACGGAGCCAT TCTGTATACA AGGTGTGCTC TTTCCAATGT 420
 AGAAGGGGTT ATGGAAGAGG GTGCGATCCT TTGCTGTAAA CTGGAGAGAC CAGTCCCAA 480
 CAGAGGGGAA TTTTAAGCCC TTCTCATCAC CCAATTGGAT GTTTTGTCTT ATAGCAAATT 540
 CCTGCAAAAT AAATAAATAA ATATTTGCAA AACTAAAAA AAAAAAAAAA AAAAAAAAAA 600
 55 GGGGGGNCNN C 611

60

(2) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 2632 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

10 TCCCAGCTCT CAGGACAAGG GCCCTGGGCG ATCTTTTAAA AAAGCCGATT GGGTGTCTTT 60
 CTAAAANTAC AACCACTACT TCATCGTCAA GTTTCCTGGA AGGGAGTCCC CTCCAGATTC 120
 15 TCATGGAGTG ACAAATCTTG ACTCTTGCTC CTGGAATTTT TCAGGCCCAA ACTAGCGTTT 180
 CTACAATGAT TTATTTGCCA AATTGTCTTT GATTATGGGT GGCTGATGAG GAACGTGCTT 240
 20 TTGTTAGGAA CCGAACTGG GCGGCGGTGA GGGCGTGAC GCAATGAGTC CGGAAGAGGG 300
 TGAAATGCTT TCGGTAGGCA CTCCACGGCT GTGAAGATGG CGGCGGCTGC GTGGCTTCAG 360
 GTGTGCGCTG TCATCTTCTT GCTTCTGGA GCTCAGCCGT CACCACTGTC GTTTTTCAGT 420
 25 GCGGGACCGG CAACCGTAGC TGCTGCGAC CGGTCCAAAT GGCACATTCG GATACCGTCG 480
 GGGAAAAATT ATTTTAGTTT TGGAAAGATC CTCTTCAGAA ATACCACTAT CTTCTGAAG 540
 TTTGATGGAG AACCTTGTA CCTGTCTTTG AATATAACCT GGTATCTGAA AAGCGCTGAT 600
 30 TGTTACAATG AAATCTATAA CTTCAAGGCA GAAGAAGTAG AGTGTATTT GGAAAACTT 660
 AAGGAAAAAA GAGGCTTGTC TGGGAAATAT CAAACATCAT CAAAATGTT CCAGAACTGC 720
 35 AGTGAACCTT TTAACACACA GACCTTTTCT GGAGATTTTA TGCATCGACT GCCTCTTTTA 780
 GGAGAAAAAC AGGAGCTAA GGAGAATGGA ACAAACTTA CCTTATATGG AGACAAAACC 840
 GCAATGCATG AACCATGCA AACTTGCAA GATGCACCAT ACATTTTAT TGTACATATT 900
 40 GGCATTTTCT CCTCAAAGGA ATCATCAAAA GAAAATTCAC TGAGTAATCT TTTTACCATG 960
 ACTGTGAAG TGAAGGTCC CTATGAATAC CTCACACTTG AAGACTATCC CTTGATGATT 1020
 45 TTTTTCATGG TGATGTGAT TGTATATGTC CTGTTTGGTG TTCTGTGGCT GGCATGGTCT 1080
 GCCTGCTACT GGAGAGATCT CCTGAGAATT CAGTTTGGGA TTGGTGCTGT CATCTTCCTG 1140
 GGAATGCTTG AGAAAGCTGT CTTCTATGCG GAATTCAGA ATATCCGATA CAAAGGARAA 1200
 50 TCTGTCCAGG GTGCTTTGAT CCTTGCAGAR CTGCTTTCAG CAGTGAACG CTCCTGGCT 1260
 CGAACCCCTG TCATCATAGT CAGTCTGGA TATGGCATCG TCAAGCCAG CCTGGAGTCA 1320
 55 CTCTTCATAA GGTGTAGTA GCAGRAGCCC TCTATCTTT GTTCTCTGGC ATGGAAGGGG 1380
 TCCTCAGAGT TACTGGGGCC CAGACTGATC TTGCTTCCTT GGCCTTTATC CCCTTGGCTT 1440
 60 TCCTAGACAC TGCTTGTGC TGGTGGATAT TTATTAGCCT GACTCAAACA ATGAAGCTAT 1500

362

TAAAACTTCG GAGGAACATT GTAAAACTCT CTTTGTATCG GCATTTCACC AACACGCCTA 1560
 TTTTGGCAGT GGCAGCATCC ATTGTGTTTA TCATCTGGAC AACCATGAAG TTCAGAATAG 1620
 5 TGACATGTCA GTCGGACTGG CGGGAGCTGT GGGTAGACGA TGCCATCTGG CGCTTGCTGT 1680
 TCTCCATGAT CCTCTTTGTC ATCATGGTTC TCTGGCGACC ATCTGCAAAC AACCAGAGGT 1740
 10 TTGCCTTTTC ACCATTGTCT GAGGAAGAGG AGGAGGATGA ACAAAGGAG CCTATGCTGA 1800
 AAGAAAGCTT TGAAGGAATG AAAATGAGAA GTACCAAACA AGAACCCAAT GGAAATAGTA 1860
 AAGTTAACA AGCACAGGAA GATGATTGA AGTGGGTAGA AGAGAATGTT CCTTCTCTG 1920
 15 TGACAGATGT AGCACTTCCA GCCCTTCTGG ATTGAGATGA GGAACGAATG ATCACACACT 1980
 TTGAAAGGTC CAAAATGGAG TAAGGAATGG GAAGATTTC AGTTAAAGAT GGCTACCATC 2040
 AGGAAGAGA TCAGCATCTG TGTCAGTCTT CTGTACGGCT CCATGGGATT AAAGGAAGCA 2100
 20 ATGACATCCT GATCTGTTC TTGATCTTTG GGCATTGGAG TTGGCGAGAG GTGTGAGAAC 2160
 AAAGAGAACA TCTTACTGAA AACAAGTCA TAAGATGAGA AAAATCTACG AGCTTCTTAT 2220
 25 TTACAACACT GCTGCCCCCT TTCTCCAG ACTCTGACAT GGATGTCAT GCAACTTAAG 2280
 TGTGTGTTTC CTGAACCTTC TGTAATGTTT CATTTTTTAA ATCTGACAAA CTAAAAAGTT 2340
 TAACGTCTTC TAAAGATTG TCATCAACAC CATAATATGT AATCTCCAGG AGCAACTGCC 2400
 30 TGTAATTTTT ATTTATTTAG GGAGTTACAT AGGTGATGGG GGAAATGTT AACTACCTTT 2460
 CATTTCTCTG GGAAGTCAAG GTTACATCTT GCAGAGGTTG TTTTGAGAAA AAAGGGCCCT 2520
 35 TCTGAGTTAA GGAGCCATAG TTCTATCAAT GATCAAAAGA AAAAAAAAAA AACTCGATCG 2580
 GCACGAGGGG GGGCCCGTA CCCAATTGCG CCTATGGGAN TCGAATGAGA CC 2632

40

(2) INFORMATION FOR SEQ ID NO: 111:

45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2249 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

GAATTCGGCA CGAGCTCACC GTGCTGCGTG ACACAAGGCC AGCCTGCGCC TACGAGCCCA 60
 TGGACTTTKT RATGGCCCTC ATCTACGACA TGGTACTGSW TGTGGTCACC CTGGGCTGG 120
 55 CCCTCTTCAC TCTGTGGGC AAGTCAAGA GGTGGAAGCT GAACGGGGCC TTCTCTCTCA 180
 TCACAGCCTT CCTCTCTGTG CTCATCTGGG TGGCCTGGAT GACCATGTAC CTCTTGGCA 240
 60 ATGTCAAGCT GCAGCAGGG GATGCCTGGA ACGACCCAC CTTGGCCATC ACGCTGGCGG 300

	CCAGCGCTGG GTCTTCGTCA TCTTCCACGC CATCCCTGAG ATCCACTGCA CCCTTCTGCC	360
5	AGCCCTGCAG GAGAACACGC CCAACTACTT CGACACGTGC CAGCCCAGGA TCGGGGAGAC	420
	GGCCTTCGAG GAGGACGTGC AGCTGCCGCG GGCCTATATG GAGAACAAGG CCTTCTCCAT	480
	GGATGAACAC AATGCAGCTC TCCGAACAGC AGGATTTCCT AACGGCAGCT TGGGAAAAAG	540
10	ACCCAGTGGC AGCTTGGGGA AAAGACCCAG CGCTCCGTTT AGAAGCAACG TGTATCAGCC	600
	AACTGAGATG GCCGTGCTGC TCAACGGTGG GACCATCCCA ACTGCTCCGC CAAGTCACAC	660
15	AGGAAGAMAC CTTTGGTGAA AGACTTTAAG TTCCAGAGAA TCAGAATTTC TCTTACCGAT	720
	TTGCCCTCCCT GCGTGTGTCT TTCCTGAGGG AGAAATCGGT AACAGTTGCC GAACCAGGCC	780
	GCCTCACAGC CAGGAAATTT GGAAATCCTA GCCAAGGGGA TTTCGTGTAA ATGTGAACAC	840
20	TGACGAACTG AAAAGCTAAC ACCGACTGCC CGCCCCCTCC CTGCCACACA CACAGACACG	900
	TAATACCAGA CCAACCTCAA TCCCCGCAAA CTAAAGCAAA GCTAATTGCA AATAGTATTA	960
25	GGCTCACTGG AAAATGTGGC TGGGAAGACT GTTTCATCCT CTGGGGGTAG AACAGAACCA	1020
	AATTACAGC TGGTGGGCCA GACTGGTGTG GTTTGGAGGT GGGGGGCTCC CACTCTTATC	1080
	ACCTCTCCCC AGCAAGTGCT GGACCCAGG TAGCCTCTTG GAGATGACCG TTGCGTTGAG	1140
30	GACAAATGGG GACTTTGCCA CCGGCTTTGC CTGGTGGTTT GCACATTTCA GGGGGGTCAG	1200
	GAGAGTTAAG GAGGTTGTGG GTGGGATTCC AAGGTGAGGC CCAACTGAAT CGTGGGGTGA	1260
35	GCTTTATAGC CAGTAGAGGT GGAGGGACCC TGGCATGTGC CAAAGAAGAG GCCCTCTGGG	1320
	TGATGAAGTG ACCATCACAT TTGGAAAGTG ATCAACCACT GTTCCTTCTA TGGGGCTCTT	1380
	GCTCTAGTGT CTATGGTGAG AACACAGGCC CCGCCCCCTC CCTTGTAGAG CCATAGAAAT	1440
40	ATTCTGGCTT GGGGCAGCAG TCCCTCTTC CCTTGATCAT CTGGCCCTGT TCCTACACTT	1500
	ACGGGTGTAT CTCCAAATC TCTCCCAATT TTATTCCTT ATTCAITTC AAGACTCCAA	1560
45	TGGGCTCTCC AGCTGAAANS COCTCCGGGA GGCAGGTTGG AAGGCAGGCA CCACGGCAGG	1620
	TTTTCCGCGA TGATGTCACC TAGCAGGGCT TCAGGGGTTT CCACTAGGAT GCAGAGATGA	1680
	CCTCTCGCTG CCTCACAAGC AGTGACACCT CGGGTCCTTT CCGTTGCTAT GGTGAAATTT	1740
50	CCTGGATGGA ATGGATCACA TGAGGGTTTC TTGTTGCTTT TGGAGGGTGT GGGGGATATT	1800
	TTGTTTGGT TTTTCTGCAG GTTCATGAA AACAGCCCTT TTCCAAGCCC ATTGTTTCTG	1860
55	TCATGGTTTC CATCTGTCT GAGCAAGTCA TTCTTTTGT ATTAGCATT TCGAACATCT	1920
	CGCCATTCA AAGCCCCCAT GTTCTCTGCA CTGTTTGGCC AGCATAACCT CTAGCATCGA	1980
	TTCAAAGCAG AGTTTAAACC TGACGGCATG GAATGTATAA ATGAGGGTGG GTCCTTCTGC	2040
60	AGATACTCTA ATCACTACAT TGCTTTTCT ATAAACTAC CCATAAGCCT TTAACCTTTA	2100

364

AAGAAAAATG AAAAAGGTTA GTGTTTGGGG GCGGGGGGAG GACTGACCGC TTCATAAGCC 2160
 AGTACGTCG AGCTGAGTAT GTTCAATAA ACCTTTGAT ATTCTCAAA AAAAAAAAAA 2220
 AAAAAACCCG GGGGGGGGCC CGGACCTGG 2249

10

(2) INFORMATION FOR SEQ ID NO: 112:

15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2193 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

25

GATACTATAA GGCAAGTAC TCACGGGTGC GCGTTAGAC TAGTGGATCC CGGGTGCAGG 60
 AATTGGGCAG AGCGCCGCG GAGCCGAGT GCTGGCGCCC CCGCGGCCG TGCCTCCGCG 120
 GANCCAAAA TCATGAATC CACCTGAGG ACCCCGAAGA AAAGGAGGAA TTCGCCGTGC 180
 CCGAGAATAG CTCCTCCAG CAGTTTAAAG AAGAAATCTC TAAACGTTTT AAATCACATA 240
 CTGAACCACT TGTTTGAAA TTGCTGAAA AAATTTGAA AGATCAAGAT ACCTTGAGTC 300
 AGCATGGAAT TCATGATGA CTACTGTTC ACCTTGTCAT TAAAACACAA AACAGGCTC 360
 AGGATCATTG AGCTCAGCA ACAATACAG CTGGAAGCAA TGTTACTACA TCATCAACTC 420
 CTAATAGTAA CTCACACTT GGTCTGCTA CTAGCAACCC TTTTGGTTTA GGTGGCCTTG 480
 GGGGACTTGC AGGCTGAGT AGCTGGGTT TGAATACTAC CAACTTCTCT GAACTACAGA 540
 GTCAGATGCA GCGACAACCT TTGCTTAACC CTGAAATGAT GGTCCAGATC ATGGAAAAC 600
 CCYTGTGTTA GAGCATGCTC CTCAAATCCT GACCTGATG AGACAGTTAA TTATGGCCAA 660
 TCCACAAATG CAGCAGTGA TACAGAGAA TCCAGAAAT TAGTCATATG TTGAATAATC 720
 CAGATATAAT GAGACAACG TTGGAACCTG CCCAGGAATC CAGCAATGAT GCAGGAGATG 780
 ATGAGGAACC AGGACCGAC TTTGAGCAAC CTAGAAAGCA TCCAGGGGG ATATAATGCT 840
 TTAAGGCGCA TGTACAGGA TATCAGGAA CCAATGCTGA GTGCTGCACA AGAGCAGTTT 900
 GGTGGTAATC CATTTGCTC CTGCTGAGC AATACATCCT CTGGTGAAGG TAGTCAACCT 960
 TCCGTACAG AAAATAGGA TCCCTACCC AATCCATGGG CTCCACAGAC TTCCAGAGT 1020
 TCATCAGCTT CCAGCGGCAC TGCTAGCACT GTGGGTGGCA CTACTGGTAG TACTGCCAGT 1080
 GGCATTCTG GGCAGAGTAC TACTGCGCCA AATTTGGTGC CTGGAGTAGG AGCTAGTATG 1140
 TTCAACACAC CAGGAATGCA GAGTTGTTG CAACAAATAA CTGAAAACCC ACAACTTATG 1200

60

365

CAAAACATGT TGTCTGCCCC CTACATGAGA AGCATGATGC AGTCACTAAG CCAGAATCCT 1260
 GACCTTGCTG CACAGATGAT GCTGAATAAT CCCCTATTTC CTGGAAATCC TCAGCTTCAA 1320
 5 GAACAAATGA GACAACAGCT CCCAACTTTC CTCCAACAAA TGCAGAATCC TGATACACTA 1380
 TCAGCAATGT CAAACCCTAG AGCAATGCAG GCCTTGTTAC AGATTGAGCA GGGTTTACAG 1440
 10 ACATTAGCAA CGGAAGCCCC GGGCCTCATC CCAGGGTTTA CTCCTGGCTT GGGGGCATT 1500
 GGAAGCACTG GAGGCTCTTC GGGAACTAAT GGATCTAAG CCACACCTAG TGAAAACACA 1560
 AGTCCCACAG CAGGAACCAC TGAACCTGGA CATCAGCAGT TTATTGAGCA GATGCTGCAG 1620
 15 GCTCTTGCTG GAGTAAATCC TCAGCTACAG AATCCAGAAG TCAGATTTC GCAACAACCTG 1680
 GAACAACTCA GTGCAATGGG ATTTTGTGAAC CGTGAAGCAA ACTTGCAAGC TCTAATAGCA 1740
 ACAGGAGGTG ATATCAATGC AGCTATTGAA AGGTTACTGG GCTCCAGCC ATCATAGCAG 1800
 20 CATTTCTGTA TCTKGAAAAA ATGTAATTTA TTTTGTATAA CGGCTCTTAA ACTTTAAAT 1860
 ACCTGCTTAA TTTTATTTTG ACTCTTGGA TTCTGTGCTG TTATAAACAA ACCCAATATG 1920
 25 ATGCATTTA AGGTGGAGTA CAGTAAGATG TGTGGGTTTT TCTGTATTTT TCTTTTCTGG 1980
 AACAGTGGGA ATTAAGGCTA CTGCATGCAT CACTTCTGCA TTTATTGTAA TTTTTTAAAA 2040
 ACATCACCTT TTATAGTTGG GTGACCAGAT TTGTGCTGC ATCTGTCCAG TTTATTGCT 2100
 30 TTTTAAACAT TAGCCTATGG TAGTAATTTA TGTAGAATAA AAGCATTAAG AAGAAGCAAA 2160
 AAAAAAAAAA AAAAATTCCT GCGCCCGCGA ATTCTTCT 2198
 35

(2) INFORMATION FOR SEQ ID NO: 113:

- 40 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1043 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 45 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

CTGAAGTGTA TGTGGTGAGG AAGAAGAGGC TCCTACTGTA GACAGCCTTG TTCTACAGAT 60
 50 CCTCCAGAA ATCTCTGGGC CAGGTGGAAC CCAGGTCAG AGAGGGATGG GAGAGAGGTT 120
 TAATTTTCCA TGATAAATAA AAATCTATAA AATAATAAAC AAGAGAAAAG AGATTGGAAA 180
 CAGCCAGGTT GGAGCAGTGA GTGAGTAAGG AAACCTGGCT GCCCTCTCCA GATTCCCCAG 240
 55 GCTCTCAGAG AAGATCAGCA GAAAGTCTGC AAGACCTAA GAACCATCAG CCTCAGCTG 300
 CACCTCCTCC CCTCCAAGGA TGACAAAGGC GCTACTCATC TATTTGGTCA GCAGCTTTCT 360
 60 TGCCCTAAAT CAGGCCAGCC TCATCAGTCG CTGTGACTTG GCCCAGGTGC TGCAGCTGGA 420

366

5 RGACTTGGAT GGGTTTGAGG GTTACTCCCT GAGTGA CTGTGCCTGG CTTTGTGGA 480
 AAGCAAGTTC AACATATCAA AGATWAATGA AAATGCAGAT GGAAGCTTTG ACTATGGSCT 540
 CTTCCAGATC AACAGCCACT ACTGGTGCAA CRATTATAAG AGTTACTCGG AAAACCTTTG 600
 CCACGTAGAC TGTCAAGATC TGCTGAATCC CAACCTTCTT GCAGGCATCC ACTGCGCAA 660
 10 AAGGATTGTG TCCGGAGCAC GGGGGATGAA CAACTGGGT AGAATGGAAG KTTGCACTGT 720
 TCAGGCCGGC CACTCTTCTA CTGGCTGACA GGATGCCGCC TGAGATKAAA CARGGTGCCG 780
 GTGCACCGTG GARTCATTC AAGACTCCTG TCCTCACTCA RGGATTCTTC ATTTCTTCTT 840
 15 CCTACTGCCT CCACMTCATG TTATTTTCTT CCCTTCCCAT TTACAATAA AACTGACCAG 900
 AGCCCCAGGA ATAAATGGTT TTCTTGGCTT CCTCCTTACT CCCATCTGGA CCCAGTCCCC 960
 20 TGGTTCCTGT CTGTTATTTG TAAACTGAGG ACCACAATAA AGAAATCTTT ATATTATCG 1020
 AAAAAAAAAA AAAAAAACT CGA 1043

25

(2) INFORMATION FOR SEQ ID NO: 114:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 703 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

GAATTGGCA CGAGTGGCG GGCACCACG CGGTTTTCG ACGCTGGCGG TGGACGCAGG 60
 CAGCATGGAC CACGGTGTCT GGGCGGATGG GGAGCGTCTA TGGTCAGTTG CCTTAGAAGT 120
 40 GGTGAGATGG GAACTGCAG TTGGAAGACC CTGGAGGATG OCTGACAAGG GGATGTCTGA 180
 CACATGATG GAGCTCTTTT TGAAATGTTT CTGCCCCTTC CTGGAGCAGA GGAGCCATTA 240
 45 TTTATGCAGG TACATCGAAG TCTTTTGACC TCATACAGT GATTATGCTT GTCATGCTG 300
 GTGGTATCCT GGCGGCCTTG CTCCTGCTGA TAGTTGTGCT GCTCTGTCTT TACTTCAAAA 360
 TACACAACGC GCTAAAAGCT GCAAAGGAAC CTGAAGCTGT GGCTGTAAAA AATCACAACC 420
 50 CAGACAAGGT GTGGTGGGCC AAGAACAGCC AGGCCAAAAC CATTGCCACG GAGTCTTGTC 480
 CTGCCCTGCA GTGCTGTGAA GGATATAGAA TGTGTGCCAG TTTTGATTCC CTGCCACCTT 540
 55 GCTGTGTGCGA CATAAATGAG GGCCTCTGAG TTAGGAAAGG TGGGCACAAA AATCTTCATG 600
 AGCAATACTT CTTAGTAGAT TGTTTTGTGA TTCAAATCAA GTTCTAGTGT TTTTATGTGA 660
 GATTATATAA TTTACAGTGT TGTTTTATAT ACTTTTGAAT AAA 703
 60

(2) INFORMATION FOR SEQ ID NO: 115:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3684 base pairs

(B) TYPE: nucleic acid

10

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

15

GGCAGAGGGG GCATGAGCAG GAGGAGGATT ACCGCTACGA GGTGCTCAGC GCCGAGCAGA 60

TTCTACAACA CATGGTGGNA ATGTATCCGG GAGGTCAACG AGGTCATCCA GAATCCAGCA 120

ACTATCACAA GAATACTCCT TAGCCACTTC AATTGGGATA AAGAGAAGCT AATGGAAGG 180

20

TACTTTGATG GAAACCTGGA GAAGCTCTTT GCTGAGTGTC ATGTAATTAA TCCAAGTAAA 240

AAGTCTCGAA CAGCCAGAT GAATACAAGG TCATCAGCAC AGGATATGCC TTGTCAGATC 300

25

TGCTACTTGA ACTACCCTAA CTCGTATTTC ACTGGCCTTG AATGTGGACA TAAGTTTGT 360

ATGCAGTGCT GGAGTGAATA TTAACTACC AAAATAATGG AAGAAGGCAT GGGTCAGACT 420

ATTTCTGTGTC CTGCTCATGG TTGTGATATC TTAGTGGATG ACAACACAGT TATGCGCCTG 480

30

ATCACAGATT CAAAAGTTAA ATTAAAGTAT CAGCATTTAA TAACAAATAG CTTTGTAGAG 540

TGCAATCGAC TGTTAAAGTG GTGTCTGCC CCAGATTGCC ACCATGTTGT TAAAGTCCAA 600

35

TATCTGATG CTAAACCTGT TCGCTGCAA TGTGGGCGCC AATTTTGCTT TAACTGTGGA 660

GAAAATTGGC ATGATCTCTG TAAATGTAAG TGTTAAAGA AATGGATTAA AAAGTGTGAT 720

GATGACAGTG AAACCTCCAA TTGGATTGCA GCCAACACAA AGGAATGTCC CAAATGCCAT 780

40

GTCACAATTG AGAAGGATGG TGGTGTAAAT CACATGGTCT GTCGTAACCA GAATTGTAAA 840

GCAGAGTTT GCTGGGTGTG TCTTGGCCCA TGGGAACCAC ATGGATCTGC CTGGTACAAC 900

45

TGTAACCGCT ATAATGAGGA TGATCCAAAG GCAGCAAGAG ATGCACAGGA GCGATCTAGG 960

GCAGCCCTGC AGAGGTACCT GTTCTACTGT AATCGCTATA TGAACCATAT GCAGAGCCTG 1020

CGCTTTGAGC ACAAACTATA TGCTCAGGTG AAACAGAAAA TGGAGGAGAT GCAGCAGCAC 1080

50

AACATGTCCT GGATTGAGGT GCAGTCCCTG AAGAAGOCAG TTGATGTCTT CTGCCAGTGT 1140

CGTGCCACAC TCATGTACAC TTATGTCTTC GCTTTCTACC TCAAAAAGAA TAACCAGTCC 1200

55

ATTATCTTTG AGAATAACCA AGCAGATCTA GAGAATGCCA CAGAGGTGCT CTCGGGCTAC 1260

CTTGAACGAG ATATTTCCCA AGATTCTCTG CAGGATATAA AGCAGAAAGT ACAAGACAAG 1320

TACAGATACT GTGAGAGTCG ACGAAGGGTT TTGTTACAGC ATGTGCATGA AGGCTATGAA 1380

60

AAAGATCTGT GGGAGTACAT TGAGGACTGA GAATGGCCCT GCATAAAATG AACTCTGAAA 1440

	ACTTTACCAT CTAGAGTGCT CATGCAATTA AAACAAAACA AACACAAACA AGGAGGCACT	1500
	AAGCCTATTC TGACACCACT GGTCTGTAGT ACCAGAATIG TTTTGTTAAT GGAAAGTTTA	1560
5	AGTAAATTAT ATTGTAATAA AAAGGTAGAT AAACCATTGT ACAACAGTAT TCTAGGCCGC	1620
	CAACAAAAGT GTGACAGACA CACTAAAAGC CCTCCAACCTT TAACTTGTAAC CGTAGCTTCA	1680
10	TTCTCAAAGC TGACTCCTTT TTTTCTTTT TCCTTTTCCT GAGTGTAGTA CAGTTAAAAT	1740
	TTCAAACAGC TCCTTGACAC TGCTTTTCAT GTTCAAACCA GCCATTTTGT TGTACTTTGG	1800
	TAAAGGACCT CTTCCCTTC CTCCCTACA CATAAGATA CACCCACACA CAGACTGACT	1860
15	CTCTTTCTCT CATACCCCAA GGTGATGAGT GAATGATGCT TAGTTCCTTG TAAAGAAAAT	1920
	CTTGGGATGG GGAAGGGGT AGGCAGCAAG AGGATTCAAC AAACGAAAAA CATAAAAACT	1980
20	TTGTATATGA CTTTAAAC AAGAGGACAA CACAGTATTT TTCAAATG TATATAGCGC	2040
	ATATGCATGG ACAAGCAAG CGTGGCACGT GTTGCATAA TGTTTAATTA CAAAAAATA	2100
	TTTATCTTT AAAATCTTC AAGATTATGT CTATTGCTG TGCATTTTCT TTCAGTTTGC	2160
25	TTATCTTCC CGGGTGGGG TTGGGATAAA GGTGTGCGG TTTAGCACCT CTGGAAGACC	2220
	TATCTAGAGC TCTTCACTT TCCTGAGGTT ATTTTGCCY TTCTGGTGT GGTATGTCTG	2280
30	TTGCCGCCA TGGCTNCAY GCCTTGAAIT OCTGCTCTG ATCAGGGACA AGGGAGGTCA	2340
	AGCTCTGACT AATGCCATGA CCTGATTAAG GGTACAGCA GGGAGTTTG TTGCTACAGC	2400
	TCATGAATTA ACCTGTCCA ACCTAATCCC CCTCCATGGC ATCATGCCTC TACCCAAGCC	2460
35	TTTGTGTGCC CATGTTATGC ACACAGCTGT AGGCATTCTT AAGTCCCTG TCGCATCCAG	2520
	TGGAAGCATT TTAAATTTT TTTTACTTTT TGGTTTCCC TTAATGCTG CTTTTCAGAT	2580
40	TTTAGTTATG GCTGCTGTC TCACCCCTTC TCTACATTAG GGTGTCAAAG AGAATGTTTT	2640
	GCTTTAAATA TAAATAGCCA TTCATTTAGT CTCAGATTGT GAATTTAAAA TGGTGGATAC	2700
	CGAAATGCT TGTGTGTGTT GCTGTGGGT TGGTTTGAAG GCAACACCC CTAGAACATG	2760
45	ATATTCCCAT CTAGTGCAAT TAAATAGAAA TCACTGAGTT TGCTGCTTTT TTATTGTCAG	2820
	CAGATAGGAG AATTAATAAT GCATTTTAGC TGTGATGTCC ATTTTATGA AATTCCTACT	2880
50	AAGAGCTATG TTAAAGTAA AGGATGGTGG TGGTTGTATT AACTATATAC CTGTTTAGGC	2940
	CATTCTGGCT GTGGTATTT TCAATAGGTC AGCATCTGTA AATCTGTCAG TTTTATACAG	3000
	GAGTGCAGAG TGAAGTAGGC AACTAGATTA AGAGGTCTAA ATATGAAATA CCAGTTGAGG	3060
55	CTGAGGACCT CTTGCTCTTC CTTTAAATGT CTTTGCCTA GGGAGTGTG ACCATTGTG	3120
	AGGCAGCTTT GTCTGCTCTT AACTGTACA TCCTATTACT CCAITGGGAA GTAGGTTTAC	3180
60	TTTCTCTGG CCTTTGCTT AAGTTAGGCT TTGCTGAATC AACCTACTT TTCTTTTAG	3240

AAAAGGTTGT TACAGGAGAT TTACTGGCAA CTGTTCTTTT CCCATCAAAA ATCAGTGAAT 3300
 GTTGTCTGAG TATAAATGCT GCTTCCTTAA ACCACTTGTC GCTTTAGGAT CAACTTTACC 3360
 5 TGTACCTTTT CTCCTTTCCT CCCTTGOCAC CTCAGGTGCA AATCTGAACT CAGTGTCTGC 3420
 TTCTTCCATT TTCTCGTCTC TCTCCCTCTT TCCCCCAFTA TCCATATGAC ATTATTTTAC 3480
 10 TTCAAATGAC AGCATCAATC TTAAAAAGAT ATACATTAAA ACTAAGGAGT TTTTTTAAAG 3540
 AAAGCCTGAA TAAGTTCCTT TCCCTGGTAA CTTTGAAAAG CAGTCAGAGT TGCTATATAG 3600
 ATATATGTGG CTCCTTTAAA ATGCTTTGTG TATGTGTGGT GTTTAAAAAA AAAAAAAAAA 3660
 15 TTCGGGGGGG GGCCCGGTTC CCAT 3684

20

(2) INFORMATION FOR SEQ ID NO: 116:

(i) SEQUENCE CHARACTERISTICS:

25

- (A) LENGTH: 1965 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

30

AAGAAAGGGT ATTAAATTC TAGATCACAT ATGGACCCGG GAAGGTTTTT NACCTCTGT 60
 TAGTGACATC GAGTCTCCCA CTAGACAAA TAGGTGGAAA AATCTCTCGA GGGCTCACAT 120
 35 TGTTTTGTC TCTTCAGGAA AACACACCACC AGGCCATACC ACAGCCTGCC CAGTGAGGCG 180
 GTCTTTGCCA ACAGCACCCG GATGCTGGTG GTGGCCTTG GGCTGCTGGT GCTCTACATC 240
 CTTCTGGCTT CATCTTGAA GCGCCAGAG CCGGGGATCC TGACCGACAG ACAGCCCCTG 300
 40 CTGCATGATG GGGAGTGAAG CAGCAGGAAG GGGCTCCAA GAGCTCCTGG TGGTGCAGCC 360
 TGTGCTCCCC TCAGAAGCTC TGCTCTTCCC AGGGCTCCCG GCTGGTTTCA GCAGGCGACT 420
 45 TTCTTCCAAT GCTGGGCCCA GACTTCTTGC CTGGGTGCTG GCCTGCCCTC TCCGNNCCGC 480
 TTGCTGCCTG TCTGCTTCC TTGGTGGYTT TGCTGGGTGC TGGGCCTGCC CTCTCCGGCC 540
 50 GCTTGTCTGC TGCTGCTTT CCTTGGTGGC TTGCTGGGT GCTGGGCCTG CCTTCTCTGG 600
 CTGCTTGCTG CCTGTCTGCT TTCCTTGGTG GCTTTGGCTT CTGCACTCCT TGGCGTCASC 660
 TCTCAGGTCC TCCATTACA CGAGTCTTC CTGCTCTGG CCGCTCTTGC TGCTCCTGTC 720
 55 TGAAGAWATC AGACTGATTT CCTCTTAAGA CTCCTAGGGA TGTGGTGAAG AGCTGGGACT 780
 CAAGTGCACT CCACGGTGTG AAACATGAGG GARGTGAGGT GTCCGTCCAC TTCCCCATA 840
 60 AAGGTGTGCA TTTCAGTTAG GCTGCCCCGC CACAGAGCAG GCTTCATCTG CTCTGCCATC 900

370

CAGCCCCATC TGGATGTGAG GTGGGGTGA GACATCATGG GGTGATTGCA GAAAGGGGA 960
 GTGGCGGCCC ACGCAGCTTC TGCTGAGGAG CTGACCGCTC TGAGCTGTTC TGTTCGTAT 1020
 5 TGCTGCTCTG TGTCTGCATG TATTGTGACC GTGGGGCTCC ACCTCTTCCA GCTGCTGCTA 1080
 CAGCTGAGGC CTGGATCCCG GCCTTTCCCT GTGACTTACG TGTCTGTAC CCGCANGCAG 1140
 10 CCTACAAAT CCTGGTGACC TGCTCTCCCA AGAACAGAGC CTGTCCCCAG ATGTCCCACT 1200
 AGCGATGAGT AACAGAGGTG GCTGTGGACT TCCTCTACTT CTCCTTGCTG GATCAGGGCC 1260
 TTCTGCCTC CCGCTGGGCA GGTCTGGCCT TGCTCTCTTG GCAGGGCCCC AGCCCCCTCTG 1320
 15 ACCACTCTGC AGCTACCAT GCAGCTGATG CCAAAGTTGT GGTGTCCAGT GTGCAGCAGC 1380
 CCTGGGAGCC ACTGCCACCT TCAGAGGGGT TCCTTGCTGA GACCCACATT GCTTCACCTG 1440
 GCCCCACCAT GGCTGCTTGC CTGGCCCAAC CTAGCGTTCT GTGCCATGCT AGAGCTTGAG 1500
 20 CTGTTGCTCT TCTTCAGGG AGGAAATAGG GTGGAGAGCG GGAAGGGTCT TGCTCCTAAG 1560
 TGTGTCTGCT GTGCCTTTT TGCCTTCTCC AAAGACGCAC TGCCAGGTCC CAAGCTTCAG 1620
 25 ACTGCTGTGC TTAGTAAGCA AGTGAGAAGC CTGGGGTTTG GAGCCACCT ACTCTCTGGC 1680
 AGCATCAGCA TCCTACTCCT GGCAACATCA GGCCAACGTC CACCCAGCC TCACATTGCC 1740
 AGATGTTGGC AGAAGGGCTA ATATTGACCG TCTTGACTGG CTGGAGCCTT CAAAGCCACT 1800
 30 GGGATGTCCT CCAGGCACCT GGGTCCCATG ACCAGCTCCC CGTCTCCATA GGGGTAGGCA 1860
 TTTCACTGGT TTATGAAGCT CGAGTTTCAT TAAATATGTT AAGAATCAA GCTGTCTTTG 1920
 35 TTCAGGCTGC TATAACAAAA ATATAATAGC CTGGGTGGCT TAAAC 1965

40 (2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 503 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

50 AGTGATCCCC TTGCTCGGC CTCCCAAAAT GCTGGAATTG TAAGCGTGGG CCTCTGCACC 60
 CGGCCTGGTC CGCAATTAA AACGCACAG CCACCATTC CTYTCCAGAA AGCACCAGCA 120
 TGCTTTTGGG AGAACCAGCC TCCTCCATGG AGGAAAGCTT GGGATCTGCC TTCCCACTG 180
 55 GGGAGGAGAG GGATCTGTGG AAAATCCTTC TGACGGACTT CCCCTCAGTG CCTGATCCAT 240
 ACTCAATAGT AGAAAAAGTA AGAAATATAC AAAGATAGCA GATACACGGA GACAGTTCCC 300
 60 CAAATAGCTG AGCGAWTAGC GCAGAAGCAA TATTGAAGAC CTAATAGCTG AGACATTTC 360

371

AGAACTGATA AAGTGCATCC AGCCACAGAT CAAGCAGCCC AGAAAATTCC AGGCAGCATC 420
AACAATAAAA TAGCCCCACA TGCACCCGTG AAAATGCAGA AGACCAAACA AAAAAGTCCG 480
GTCAACAGCC AGAGTTAAAG AGG 503

(2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1133 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

GGCACAGCTT GGAATGAACC CCTGTGGATA AGGGGGACTA TTAGATAGAA TAAACATCAA 60
TAAATGCTTG ATGAATAAAC GCTAATCCTA CCTTCCCAGC CTGACACCTC CCAGTGGACA 120
CCACACTTCA CTTGAAGCCT TAGAAACCTT TCCCACCCAT GCTTCCAGCC CTGGCTTCAT 180
GTTGCCATTT CTCACCCCCA GAACAGGCCG CCCGCCTGAA GAACTACAA GAGCAAGAGA 240
AACAACAGAA AGTGGAGTTT CGTAAAAGGA TGGAGAAGGA GGTGTCAGAT TTCATTCAAG 300
ACAGTGGGCA GATCAAGAAA AAGTTTCAGC CAATGAACAA GATCGAGAGG AGCATACTAC 360
ATGATGTGGT GGAAGTGGCT GGCCTGACAT CCTTCTCCTT TGGGGAAGAT GATGACTGTC 420
GCTATGTCAT GATCTTCAAA AAGGAGTTTG CACCCTCAGA TGAAGAGCTA GACTCTTACC 480
GTCGTGGAGA GGAATGGGAC CCCAGAAGG CTGAGGAGAA GCGGAACNTG AAGGAGCTGG 540
CCCAGAGGCA ANGAGGAGGA GGCAGCCAG CAGGGGCCTG TGGTGGTGAG CCCTGCCAGC 600
GACTACAAGG ACAAGTACAG CCACCTCATC GGCAAGGGAG CAGCCAAAGA CGCAGCCCAC 660
ATGCTACAGG CCAATAAGAC CTACGGCTGT KTGCCCGTGG CCAATAAGAG GGACACACGC 720
TCCATTGAAG AGGCTATGAA TGAGATCAGA GCCAAGAAGC GTCTGCGGCA GAGTGGGGAA 780
GAGTGGCCGC CAACCTCCTA GCGCCCCGC CCAGCTCCCT TTGACCCCTG GGCAGGGCA 840
GGGGCAGGG AGAGACAAGG CTGCTGTAT TAGAGCCAT CCTGGAGCCC CACCTCTGAA 900
CCACCTCCTA CCAGCTGTCC CTCAGGCTGG GGGAAACAG GTGTTTGATT TGTACCGTT 960
GGAGCTTGA TATGTGCGTG GCATGTGTGT GTGTGTGTGA GAGTGTGAAT GCACAGGTGG 1020
GTATTTAATC TGTATTATTC CCCGTTCTTG GAATTTTCTT CCCATGGGGC TGGGGTACTT 1080
TACATTCAAT AAATACTGTT TAACCCAAAA AAAAAAAAAA AAAAGAAAGA AGN 1133

(2) INFORMATION FOR SEQ ID NO: 119:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1101 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

GGGCACAGCT GAAGCTGCAG ACCTCCCCAG GGGATGGCTC CTCTCCCCCA GGAGCCCCGA 60
GGCAGGGGAG GCAGAAAGCC TGGGCTCTGG GGGGTGGCCT GCGGACAGCT GTGCTGTGGG 120
CCGGGGGCTG GGCCTGTCCC ACAGGGNCGT GGAGCTCGTG GTTCTGAGCA GCCAGCTGGG 180
TGGTGTCTGG GGATAGCTGG GAGGCACAGC GGCTGCCATG TGGGACTGGG ACTGGAGTGC 240
TCCCTGGTCT TGGCCTCTGT GGCTCAGCCT TGCTCTGGTC TGCTGAGTGC CAGGGGCCAA 300
GGGGCACAGG GCCAGTGAGG CCGGCCACGC TGGGGCCCTC ACCTGTGAGA TGGGTGCGGA 360
ATTTKACACA GCCTANGGCT TGGTTCCTGG TKGINGAMCG TGGACTYCTK AGAACGGGAG 420
TGCTGGTCTT .GAAAGGCGTG GTTGAGACC AGCTGCTTTT CTGCTGTTT TTCTCTTAGG 480
AGATTAAACA AAAACAGAAA GCACAAGACG AACTCAGTAG CAGACCCAG ACTCTCCCT 540
TGCCAGAGCT GGTTCACAGC GGGGAGACGC ACCTGTGCCA GAACGGGATT CAGCTGCTCA 600
ACGGGCATGC GCCGGGGGCC GTCCCAAAACC TGCAGGGCT CCAGCAGGCC AACCGGCACC 660
ACGGACTCCT GGGTGGCGCC CTGGCGAACT TGTTTGTGAT AGTTGGGTTT GCAGCCTTTG 720
CTTACACGGT CAAGTACGTG CTGAGGAGCA TCGCGCAGGA GTGAGGCCCA GCGCCGAGA 780
CCCAAGGCGC CACTGAGGGC ACCGCGCACC AGAGCGTGAC CTGGCAGGC TGGACACACT 840
GCCCAGCACA GGCAGACCCA CCAGGCTCCT AGGTTTAGCT TTTAAAAACC TGAAAGGGGA 900
AGCAAAAACC AAAATGTGTG ACTGGGCTTT GGAGGAGACT GGAGCCTCAG CCCTGTCTCTG 960
GCCACGGGCC GCTGGGGCTG GTGTGGGTGG GCCTTGTGTG CTGGATTGT AGCTTATCTT 1020
CCGTGTGTG TTTGACCTG TTTTAGTAAA CCCGTTTTTC ATTTTAAAAA AAAAAAAAAA 1080
AAACTTTGGG GGGGGGCCCC N 1101

(2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 282 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

AGCTTCTCTG TCCAGTCTTG AACTCTGGGS TCTCTTGGAA CTTTCTTCAC CCCTCTCAGC 60
 5 CTGAATATTC CTTCCATGGA TTCCACTCAA CCAGACTTTG GATCTGTGCC TACTTAATCA 120
 ACCTTATCTT TGCAATATGT TGGGGCCAC CTTCACATCC TTGGTTCTTG TTCCTCCTTG 180
 10 GCCTAACTTG TCCCTTCTCC ACTTCACATC CCCGGTGGGA CAGCATTCCT CCTTCCTCCC 240
 AACCTCCCTC CGTCTCARAA AAAAAAAAAA AAAAAAAAAA TT 282

(2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2635 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

25 TAAGGGGGTG TGTGCTCACC TCCTCCTGAC CCTTAACACT CCTGTCTGTC CCAGACCAAC 60
 AGAGAGAGCT GTCCTGAGA CCCCGGAGAG AAGCAGCTGC CGAAAGCTGC AGCCTTTCCG 120
 30 CACTCTGAGA CATGATCTT CCTCTGCCA GGGGAGAGCC ACCCACAGGC CATGTCCAGC 180
 CCCACTTCCC TCAGCCCCCA GGGYTTCCTT CTGGCCCCTC TGAGGATTCC CTAGGGCTGC 240
 35 CCCGCAGAGG GGYTTCCCCA AGCTCTGTTT TGAAGCCTGC AATGTGGAAA AGTGAGAAGT 300
 CAGAGGGAAC AGGACAGGTG CAGCCGGGCT CTGAGGCCAC ACCTCACACC TCGCTGTTC 360
 CCAACATCCC CTGAGCAGTG TGAGCTCATC TCACCAGATG AGAAGAGGCC CTGTGCATTT 420
 40 YTTTGTGTTG TTTGTGCTG TTTTCCCCA CCCATCCAGT TCTCCTCAGC AAAGCAAATT 480
 CCTTAACACC TTTGGTGGAG AATTCTTTAC CCAGACTTGG GGCTGTGATG CCCTTCAGTG 540
 45 CGTGGTGAGT GCAGCGTGTG TGGGTGTGCC TGTGTGTGAA CCTGGGGGCC ATCCTGGTGG 600
 CCTGGGAGCG TGAGGAGAGG CCCCTGTGT GCTGGGTGAG TGGTGGGTGT GGGGTCAATG 660
 CAGTGAGGCT CTCGGGTGA GGCTCCCAAC CTGGCAGTCC CCAGCCTCCC AGCATCTGTG 720
 50 AGCGTCTGTT GGACTTTACA GAAGAGCCTC ATCCYGTCTG CCCCTCACTC TGCCCTGGAA 780
 TCAACATCTT CCGAGTCTT CTTGGGGGAA ATAGCAGAGC CCCACTTAAC TCCATAAACT 840
 55 GCTTCCCAIT CCGCAGCCCA GTTCTGATTG TTGAGGTGTC GCGTCGTTC AGGTCCCCCA 900
 GTCCCTCTT TCTCCTGTCC TCTCTGTGTC CTTACCTCC CCACTCCAGC CCCGGCTCAG 960
 TTCAGGAAA TGCTGTCCA YATCAGCCCT CTGCTCTCTG AGGCAGCCGC GCCTCTGACT 1020
 60 CGGAGCTACT TGAAACTTCT GCTCTTGCTA GGATTGGAGT CTACCTATCT CTTCCATTG 1080

	TCCCAGCTGG AGTTCITGGAA CTTTCCTCCT CGGGGTG3GG GTGGGGGTTG TTAAGGATGC	1140
5	TGGGGGGCCT GGGGAAGGAA GGAGTTCAGA GGAAGGGTGT CCCCTGTCTT CTTGATGTCA	1200
	CCCTCCGCTC CTGGGACAG TGCTCTCTCT GTCTCTGGGT CTTCTGGCTG TGCACGTTTG	1260
	TGTGTCTTTG TAAATATGTT TTAGGAAGAA AGCAAAAGG ACTGAAGTAG COTCTGGTAG	1320
10	GATTGCAGGG GTCCAGCCTT GCCTGTTTCC GAAGCCCCCA CACTGCTTTT CGCCCCACTG	1380
	AGACTGGTCC CCTCAAAAGG TAGACAAAAC AGCAGCTCCC TGTGGAGCTG AAGGGCGGCC	1440
15	TCAAAGTGGC TTTTGTGTAG ACAAGGTTAA GGTTCCTCA TGAGCAAGT TGTGATCGG	1500
	TCCTTCCTCA GCTCTTGAT TTGTGACCTT GACCAAGGGG CCTGCCCTCC AGCTCCTCCA	1560
	GTGCCCCTC CTGATGCCT CGCTCCTTCC TGCCCCACT CCCCTGGTTT AGGAGGTAG	1620
20	GGGAATTAGG GCCATGCTGG AAGAAGCTTA ACCATGTGTT CAAAGAGGG TTTCTTGCTT	1680
	GCTTGGTCTT GGAATCCTCC TTGGCTGCC CAGGCTCTT TGGCCCTGG GTGCTGGGG	1740
25	AGGTGGATGT CAGATCTGTT AGGTTCAGC AGAGAAATA AATGTGCTTT GAGGAGCCAC	1800
	TCAGAGAGGG TCCAAGGGTG ATGGAGAAGG AAGCATGCTC TGGGAGCTTG GAGGGGARGG	1860
	GTGGTGGGTG GGGCATCTT GACTGCCCCC TGTGTCTCCA CAGTGGGGG GTGGTCACCC	1920
30	CYCTTCACTC CAGCCCGCTT GCCTTCAGCC TTCCATGAGC TTCACTGCTT TCCACTTCA	1980
	CTTTGGAGGG GGTGGGGTCC GTTGGCATCA ACACGGGGAC CCTCTGCTTC ACCAAGGCC	2040
35	GAGCCCTCAG CCCCTGGGGA GAACAAATGG CTGAGCTTTG ATACCTGGGG TCTTCGAGAG	2100
	GCTGCGGGCT GGGGGCAGTC CCAGGGGAGA GACCCACAG AAGGAGCTCC AGGCTCCTCG	2160
	AGGAAGTTC CAGCAGAGCA AACTGCTTTC CAGCCTGAG CCTGCTTAAA CTGCTGTATG	2220
40	TGCAATAACT GAGCTTAGAG TTAGGAATTG TGTTCAGTG CTTGGATTTC CTTCTGTAGA	2280
	TTTAAGTCTT GAAATGTAT CTCTCAGTAA TTGTAGTGT CTTTAAAAA ATTGAAAAAC	2340
45	AAAGTGTAG ACTGTGTGCG TGTGCGTTGA TGGGCACTCA AGAGTCTCTT GAGTCATCCA	2400
	GCCCTGCTTT TCCCTGCGC CCCATCCTC TCAGTCCCG CCCGCTCTC ACTGGGGAC	2460
	CCTGCTCTGT GTGCTCTTA TCTGCTATT ACTCAGCTA AGGAAACAG TCACTCCAC	2520
50	ACATGCATAA AGGAAATCAA ATGTTATTTT TAAGAAAATG GAAATAAAA ACTTTATAAA	2580
	CACCAAAAAA AAAAAAAAAA ACCCNGGGGG GGGGCCGTA ACCCATTTGG CTTAA	2635

55

(2) INFORMATION FOR SEQ ID NO: 122:

60

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 994 base pairs

375

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

GAATTCGGCA GAGGTTCCGC GAAGATAGGG AATAAGGAAG CACAGGAGTA GGGGAGAAGG 60
 AAGCACAGGA GTAGGGGAGA TATACAGCGG TCAGGATAAG GGGGAAAGG CGGTGGTTC 120
 10 SCAAGAGGTG AAACAAGATG TGAGAGACAA GGGGTAGGGA AGAAATGGG CAGCGGTTAG 180
 GTTCAGAAGC GCATAGACCG TGGCGGACGG GCAATGCGAG GGGCACAGAA AGGAACTGAG 240
 15 GGGTGGGCTA TTTTAARGGA GATGGTCCTT CAGCCCTCTT YTTTTCTGCG TAGTTCCTCT 300
 CCTCCAGGCC GCGCGCGGAT ATGTCGTCCG GAAACCAGCC CAGTCTAGGC TGGATGATGA 360
 CCCACCTCCT TCTACGCTGC TCAAAGACTA CCAGAATGTC CCTGGAATTG AGAAGGTTGA 420
 20 TGATGTCTG AAAAGACTCT TGTCTTTGGA AATGGCCAAC AAGAAGGAGA TGCTAAAAAT 480
 CAAGCAAGAA CAGTTTATGA AGAAGATTGT TGCAAACCCA GAGGACACCA GATCCCTGGA 540
 25 GGCTCGAATT ATTGCCTTGT CTGTCAAGAT CCGCAGTTAT GAAGAACACT TGGAGAAACA 600
 TCGAAAGGAC AAAGCCCAAC AACGCTATCT GCTAATGAGC ATTGACCAGA GGAAAAAGAT 660
 GCTCAAAAAC CTCGTAACA CCAACTATGA TGTCTTTGAG AAGATATGCT GGGGGCTGGG 720
 30 AATTGAGTAC ACCTTCCCC CTCTGTATTA CCGAAGAGCC CACCGCGAT TCGTGACCAA 780
 GAAGGCTCTG TGCAATCGG TTTTCAGGA GACTCAAAAG CTGAAGAAGC GAAGAAGAGC 840
 35 CTTAAGGCT GCAGCAGCAG CCCAAAACA AGCAAAGCGG AGGAACCCAG ACAGCCCTGC 900
 CAAAGCCATA CCAAGACAC TCAAAGACAG CCAATAAATT CTGTTCAATC ATTTAAAAA 960
 40 AAAAAAAAAA AAAAAAAAAA AAAAAGGGGA GGGG 994

45 (2) INFORMATION FOR SEQ ID NO: 123:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1542 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

GGCASAGCCA CCTCGGCCCC GGGCTCCGAA GCGGCTCGG GCGGCCCTT CGGTCAACAT 60
 55 CGTAGTCCAC CCCCTCCCA TCCCAGCCC CCGGGATTC AGGCTCGCA GCGCCAGCC 120
 AGGGAGCCGG CCGGAAGCG CGATGGGGC CCCAGCCGC TCGCTCCTGC TCCTGCTCT 180
 60 GCTGTTCGCC TGCTGCTGG GCGCCGCGG GGCCAACCTC TCCCAGGAG ACAGCCAGCC 240

376

	CTGGACATCT GATGAAACAG TGGTGGCTGG TGGCACCGTG GTGCTCAAGT GCCAAGTGAA	300
5	AGATCACGAG GACTCATCCC TGCAATGGTC TTAACCCCTGC TCAGCAGACT CTCTACTTTG	360
	GGGAGAAGAG AGCCCTTCGA GATAATCGAA TTCAGCTGGT TAMCTCTACG CCCCACGAGC	420
	TCAGCATCAG CATCAGCAAT GTGGCCCTGG CAGACGAGGG CGAGTACACC TGCTCAATCT	480
10	TCACTATGCC TGTGCGAACT GCCAAGTCCC TCGTCACTGT GCTAGGAATT CCACAGAAGC	540
	CCATCATCAC TGGTTATAAA TCTTCATTAC GGGAAAAGA CACAGCCACC CTAAACTGTC	600
15	AGTCTTCTGG GAGCAAGCCT GCAGCCCGGC TCACCTGGAG AAAGGGTGAC CAAGAACTCC	660
	ACGGAGAACC AACCCGCATA CAGGAAGATC CCAATGGTAA AACCTTCACT GTCAGCAGCT	720
	CGGTGACATT CCAGTTTACC CGGGAGGATG ATGGGGCGAG CATCGTGTGC TCTGTGAACC	780
20	ATGAATCTCT AAAGGGAGCT GACAGATCCA CCTCTCAACG CATTGAAGTT TTATACACAC	840
	CAACTGCGAT GATTAGGCCA GACCCCTCCC ATCTCTGTA GGGCCAGAAG CTGTGTCTAC	900
25	ACTGTGAGGG TCGCGCAAT CCAGTCCCC AGCAGTACCT ATGGGAGAAG GAGGGCAGTG	960
	TGCCACCCCT GAAGATGACC CAGGAGAGTG CCCTGATCTT CCCTTTCCTC AACAAGAGTG	1020
	ACAGTGGCAC CTACGGCTGC ACAGCCACCA GCAACATGGG CAGCTACAAG GCCTACTACA	1080
30	CCCTCAATGT TAATGACCCC AGTCCGGTGC CCTCTCTCTC CAGCACCTAC CACGCCATCA	1140
	TCGGTGGGAT CGTGGCTTTC ATTGTCTTCC TGCTGCTCAT CATGCTCATC TTCCTTGGCC	1200
35	ACTACTTGAT CCGGCACAAA GGAACCTACC TGACACATGA GGCAAAAGGC TCCGACGATG	1260
	CTCCAGACGC GGACACGGCC ATCATCAATG CAGAAGGCGG GCAGTCAGGA GGGGACGACA	1320
	AGAAGGAATA TTTTCATCTAG AGGCGCCTGC CCACTTCTCTG CGCCCCCAG GGCCCTGTGG	1380
40	GGACTTGCTG GGGCGTCCAC CAACCGGAC TTGTACAGAG CAACCGCAGG GGCGSCCCT	1440
	CCCGMTGTGTT CCCAGCCCA CCCACCCCTT TGTTACAGAA TGTYTKGTTT GGGGTGCGGT	1500
45	TTGTWATTG GTTTNGGATN GGGGAAGGGA GGGANGCGG GG	1542

(2) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1390 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

CAAGCTCTAA TACGACTCAC TATAGGGAAA GCTGGTACGC CTGCAGGTAC CGGTCCGAA

60

377

	TTCCCGGGTC GACCCACGCG TCGGGGCTC AGGGTGGACG CATGGTTCCTG CACTGAGGCC	120
	CTCGTCATGG TGGCGCCTGT GTGGTACTTG GTAGCGGCGG CTCTGCTAGT CGGCTTTATC	180
5	CTCTTCCTGA CTCGCAGCCG GGGCCGGGCG GCATCAGCCG GCCAAGAGCC ACTGCACAAT	240
	GAGGAGCTGG CAGGAGCAGG CCGGGTGGCC CAGCCTGGGC CCCTGGAGCC TGAAGAGCCG	300
10	AGAGCTGGAG GCAGGCCTCG GCGCCGGAGG GACCTGGGCA GCGCCCTACA GGCCACAGCT	360
	CGAGCCACG GGGTGGCCTG GGCAGAAGCA GATGAGAAGC AGGAGGAAGC TGTTCATCTA	420
	GCCCAGGAGG AGGAAGGTGT CGAGAAGCCA GCGGAAATC ACCTGTGGG GAAAATTGGA	480
15	GCTAAGAAAC TGCGGAANNY GGAGGAGAAA CAAGCGCGAA AGGCCAGCK TGAGGCAGAG	540
	GAGGCTGAAC GTGARGWCG GAAACGACTC GAGTCCAGC GCGAATGAGT GGAAGAAGGA	600
20	GGAGGAGCGG CTTCGCCTGG AGGAGGAGCA GAAGGAGGAG GAGGAGAGGA AGGCCCGCGA	660
	GGAGCAGGCC CAGCGGAGC ATGAGGAGTA CCTGAACTG AAGGAGGCCT TTGTGGTGA	720
	GGAGGAAGGC GTAGGAGAGA CCATGACTGA GGAACAGTCC CAGAGCTTCC TGACAGAGTT	780
25	CATCAACTAC ATCAAGCAGT CCAAGTTGT GCTCTTGGAA GACCTGGCTT CCCAGGTGGG	840
	CCTACGCACT CAGGACACCA TAAATCGCAT CCAGGACCTG CTGGCTGAGG GGACTATAAC	900
30	AGGTGTGATT GACGACCGGG GCAAGTTCAT CTACATAACC CCAGAGGAAC TGGCCGCCGT	960
	GGCCAACCTC ATCCGACAGC GGGGCCGGT GTCCATCGCC GAGCTTGCCC AAGCCAGCAA	1020
	CTCCCTCATC GCCTGGGGCC GGGAGTCCCC TGCCCAAGCC CCAGCCTGAC CCCAGTCCTT	1080
35	CCCTCTTGA CTCAGAGTTG GTGTGGCCTA CCTGGCTATA CATCTTCATC CCTCCCCACC	1140
	ATCCTGGGGA AGTGATGGTG TGGCAGGCA GTTATAGATT AAAGGCCTGT GAGTACTGCT	1200
40	GAGCTTGGTG TGGCTTGGTG TGCAGAAGG CCTGGCCTAG GATCCTAGAT AAGCAGGTGA	1260
	AATTTAGGCT TCAGAATATA TCCGAGAGGT GGGGAGGTC CCTTGAAGC TGGTGAAGTC	1320
	CTGTTCTTAT TATGAATCCA TTCATTCAAG AAAATAGCCT GTTGCAAAAA AAAAAAAAAA	1380
45	AAAAACTCGA	1390

50 (2) INFORMATION FOR SEQ ID NO: 125:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1288 base pairs

(B) TYPE: nucleic acid

55 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

60 GCGCGCGGGG TGAAAGGCGC ATTGATGCAG CCTCGGCGG CCTCGGAGCG CGGCGGASCA 60

378

5 GACGCTGACC ACGTTCCTCT CCTCGGTCTC CTCGCGCTCC AGCTCCGCGC TGCCCGGCAG 120
 CCGGGAGCCA TGCGACCCCA GGGCCCGGCC GCCTCCCCGC AGCGGCTCCG CCGCTCCTG 180
 CTGCTCCTGC TGCTGCAGCT GCCCGCGCCG TCGAGCGCCT CTGAGATCCC CAAGGGGAAG 240
 CAAAAGGCGC ATCCGCGAGA GGGAGGTGGT GGACCTGTAT AATGGAATGT GCTTACAAGG 300
 10 GCCAGCAGGA GTGCCTGGTC GAGACGGGAG CCCTGGGGCC AATGGCATTC CGGGTACACC 360
 TGGGATCCCA GGTCCGGATG GATTCAAAGG AGAAAAGGGG GAATGTCTGA GGGAAAGCTT 420
 TGAGGAGTCC TGGACACCCA ACTACAAGCA GTGTTTCATGG AGTTCATTGA ATTATGGCAT 480
 15 AGATCTTGGG AAAATTGCGG AGTGTACATT TACAAAGATG CGTTCAAATA GTGCTCTAAG 540
 AGTTTTGTTC AGTGGCTCAC TTCGGCTAAA ATGCAGAAAT GCATGCTGTC AGCGTTGGTA 600
 20 TTTCACATTC AATGGAGCTG AATGTTTCAGG ACCTCTTCCC ATTGAAGCTA TAATTTATTT 660
 GGACCAAGGA AGCCCTGAAA TGAATTCAAC AATTAATATT CATCGCATT CTTCGTGGA 720
 AGGACTTTGT GAAGGAATTG GTGCTGGATT AGTGGATGTT GCTATCTGGG TTGGCACTTG 780
 25 TTCAGATTAC CCAAAGGAG ATGCTTCTAC TGGATGGAAT TCAGTTCTC GCATCATTAT 840
 TGAAGAACTA CCAAATAAA TGCTTTAATT TTCATTTGCT ACCTCTTTT TTATTATGCC 900
 30 TTGGAATGGT TCACTTAAAT GACATTTTAA ATAAGTTTAT GTATACATCT GAATGAAAAG 960
 CAAAGCTAAA TATGTTTACA GACCAAAGTG TGATTTTACA TGTTTTTAAA TCTAGCATT 1020
 TTCATTTTGC TTCAATCAAA AGTGGTTTCA ATATTTTTTT TAGTTGGTTA GAATACTTTC 1080
 35 TTCATAGTCA CATCTCTCA ACCTATAATT TGGGAATATT GTGTGGTCT TTTGTTTTTT 1140
 CTCTTAGTAT AGCATTTTAA AAAAAATATA AAAGCTACCA ATCTTTGTAC AATTTGTAAA 1200
 40 TGTTAAGAAT TTTTTTTATA TCTGTTAAAT AAAAATTATT TCCMACAACC TTAACAAAAA 1260
 AAAAAAAAAA AAAAAAAAAA AAAAAA 1288

45

(2) INFORMATION FOR SEQ ID NO: 126:

50

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1517 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

60

AGTGGCTTAA AGGCATCGTT TTAGGGATTA CTGGGAAGTA TCTTCAAAGT AATACATGAG 60
 AAACATTCCT TCCTAAATCC TTTATTATAT TGAATATCGT ATTAATTGGT TTCAGAGGT 120

379

TAAATTAACC ATGTATTCCT GCAATAAATG TCACTTGINT CTTGTATATA ATCTTTTTTA 180
TATATTACCG GATTGATCA TTAGTATTTT GTTGAGGATT TTTGTGTCTA TATTCATAAG 240
5 AGATGCTGGT CTGCAGTTTT CTTTTTTTGT GATAATCTGG TTTTGTATC AGTAATACAG 300
GCCCCATGAA ACGAGTTGGG AAGTGTTTAC CTCTCTTGTA TTTTTTCAAG AGTTTGTGAA 360
10 GAATGCTAT TAATCTTTA AATGTTTGGT AGAATCTACC ATTGAAATCA TGTGCTCTGG 420
GCTTTTTTTT GAGGGAAGTG TTCTGATAAC TAATTCAGTA TCTACTTTTT ATAGCTCTGT 480
TCAGATTTTG CTCTCTCTG AGTTAGTTTT GGTAATTTGT GTATCTCTAG GATTTGTGCC 540
15 ATTTCAATTA TCTCATTTGT TGGCATAAAT TAACTAAAT TTGGCCTGAG CCTACCTGTA 600
TATCTTGAGT CCCTCTGTAA GGAAGTGTAG CCTAAGTTGT ACATAAACAA ACTGAAATCC 660
TAAATTAGGA ATGTAGTTTT TGTAACAGCT CCTGAGTCTC AGGCAGTCAC AGCAGYCAAG 720
20 TCTGTCAATT GCAGGCTGCT AACTAAGCAG CCCATGSTCA AATGAGGCAA AAACCTTTGC 780
TTTTAACACA TAGTATAGCT TTGTAATCCT TTCTTGCAC ACTCGGGTAA TTCTTCCTT 840
25 TTTCATCCC KGWATTTTCC AKGAATATGA RTCTYCCTTT TTTCCCTCC TGTGAGTCTA 900
GCTAATGGTT TGTCAATTTT GTTGATCTTT TGAARAACAA ACCTTTGGTT CCACCTTTCTT 960
GTTGCATATG CTGARTATTC TCATAATTGG AGTGGAAAGC TGATCTTTGA TTACTTATTT 1020
30 TACTTAGGGC TGAGGAGTTC ATGGACTTCG CAAAACCTCC TTGAATCTAA ATTGCATCTT 1080
CTTTCCGGT TTCTGGGCTG AAACATGTTT TTTCCCATCT WANAWACCCT TGGTCTTTTC 1140
35 ATKGGCGATT AAGACTAGAG AAAGTCTAG ATMCCTTGT CTTTTATGCT GTCATTTTGT 1200
TTAAAGGCTT TCTATGTAGT AAACTATCT ATATAGACAA AATAGAGCCT TGAGTTGTGG 1260
TCTTGAATTT GATCAACATG ATTTACCACA TTCTGTACTG GATATTTCTT CACCTGCTGC 1320
40 TACTGTAAAC CATTTTATTC TTGGATCTTC TGTAGAGTAT ATTATCACAG GTACTTTTTA 1380
CAGGGGTGTC TAATCTTTTG GCTTCCCTGG GCACATTGAA AGAAGAAGAA TTGTCTTGGG 1440
45 CCACACATCA AATACGCTAA CACTAATAAT AGTTGATGAG CTAAAAAAA AAAAAAAG 1500
GCAAAAAAGN CCAAAA 1517

50

(2) INFORMATION FOR SEQ ID NO: 127:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1073 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

380

TGAATCTATT CTTTGAACAT TCTACAACAA GAATTACATT ATACTGTTAT ACCAGAGTAC 60
 TTCTGCAGTG TGAAATAGAT TGGTTTGGAA AATGAACCTG GCTTTGCTAT AAATTACATT 120
 5 CACAGGCCCTT TTTGCAAATG TGTAACCTGC CTATCAAAGT AGTTTGTAGG GCAAATGCAG 180
 AATATATGTC TCCATCTGGT AAAGTACCTT WTAYTCATGT GGGAAATCAA GTAGTATCAG 240
 10 AACTTGGTCC AATAGTCCAA TTTGTAAAAG CCAAGGGCCA TTCTCTTAGT GATGGGCTGG 300
 AGGAAGTCCA AAAAGCAGAA ATGAAAGCTT ACATGGAATT AGTCAACAAT ATGCTGTTGA 360
 CTGCAGAGCT GTATCTTCAG TGGTGTGATG AAGCTACAGT AGGGRMGATC ACTCATGMTA 420
 15 GGTATGGWTC TCCTTACCCT TGGCCTCTGW WTCATATTTT GGCCTATCAA AAACAGTGGG 480
 AAGTCAAACG TAAGNTGAAA GCTATGGAT GGGGAAAGAA GACTCTGGAC CAGGTCTTAG 540
 20 AGGATGTAGA CCAGTGTGT CAAGCTCTCT CTCAAAGACT GGAACACAA CCGTATTTCT 600
 TCAATAAGCA GCCTACTGAA CTTGACGCAC TGGTATTTGG CCATCTATAC ACCATTCTTA 660
 CCACACAATT GACAAATGAT GAACCTTCTG AGAAGGTGAA AACTATAGC AACCTCCTTG 720
 25 CTTTCTGTAG GAGAATTGAA CAGCACTATT TTGAAGATCG TGGTAAAGGC AGGCTGTGAT 780
 AGAGTTATGT GTTAGTCTCA GGAGTCTTAA CTTTTGAAAT ATGTTTACT TGAATGTTAC 840
 30 ATTAGATATT GGTGTCAGAA TTTTAAACC AAATTACTGC TTTTGAAC CTCAAATTAT 900
 ATAATGTATC TTAGTATGT GCTTTATATT GTTATTTGTG TATACATTAA AATAATTCTG 960
 AATTATTTAA TCTGATATGT TGTATTCTGT ATCTTGAAAT TTTTGTTCCT TTGAAACATG 1020
 35 CATGCATTTA AAAATAAAGC TTAAACAACT GTAAAAAAA AAAAAAAAAA CTC 1073

40

(2) INFORMATION FOR SEQ ID NO: 128:

45

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 300 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

60

CAACCCCTGC CTTTTTTTG TTTTCCATTT GCTTGGTAGA TCTTCCTCCA TCCCTTTATT 60
 TTGAGCCTAT GTGTGTCTCT GCCCGTGAGA TGAGTCTCCT GAATACAGCA CACTTACTGG 120
 55 TCTTGACTCT GTATCCAATT TGCCAGTCTG TGTCTTTCAT TTGAGCATT TAGCCCATTT 180
 ACATTTAAGG TKAATATTGT TATGTGTGAA TTTRATCYTR TCATTATGWT GTTAGCTGGT 240
 TATTTTGCTT GTTAGTTGAT GCAGTTTCTT CCNGGCATCA ATGGTCTTTA CAANTTGGCA 300

(2) INFORMATION FOR SEQ ID NO: 129:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1275 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

10

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

15 GGCAGAGCCT GTCCCTGCTG CCCCTGCAAA AAAAACCCC TCTGGTGTGA GCAGGATGGT 60
TGGAGGTAT GTGAGCTCCT TCTCCTTTCC TCCAGTTTCC TCTTCCCTTC TCCTCCCTGC 120
CTCTTTTGCT TTTCCCTTTC TTCCTGGTAC CCCCTGCCCA TTCCTGTATT TTCTCCCATC 180
20 GCCATTCTCC CCTCTCCAC TGTCCCTAAC CCGTTCAAAC TCTTTCTCT TAAATGGTTG 240
AGATTTTCTC TCACCAAGCA CACCCAGTA TTAATTAAAC TAGCTGCAA CAGGCAGCAA 300
GTGGTCTACC ATGACAGATG GGTTTTGTGT GTGTGTGTGT GTGTGTAAT GTAATAAAAC 360
25 ATATTGARTC ACTCAATAAA CACAGAGTGT CTA CTACTACATG TATCARGCAC TATCATAGAT 420
GCTAATTAAC GAAACTGAAA TGGCCAGGCC CTCACAGTGG CTCATGCCTA TAATCCCAGC 480
30 ACTTTGGGAG GATGAGGCAG GAGGATCACT TGAGGCCGGG AGTTCAAGAC CAGCCTGGGC 540
AACATAGTAA GACTCCATCT CTACAAAAA AAAATTTTTT TTATTATACT TTAAGTTTGT 600
GGTTACATGT GCAGAACGTG TAGTTTGTGT ACATAGGTAT ATACGTGCCC TGGTAGTTTG 660
35 CTGCACCCAT CAACCCATCA CCTACATTAG GTATTTCTCC TAATGTTACC CCTCTCCTAG 720
CCCCCACCC CGTGACAGGC CTTGGTGTGT GATGTTCCCC TCCCTGTGTC CATGTGTTCT 780
40 CATGGTCAA CTCTCACCTA TGGAGTGAGA ACATGTGGTA TTGGTTTTC TGATCTTGIG 840
ATAGCTTGCT GAGAAATGKG GTTCCAGCT TTATCCACGT CCTGCAAAG GGCATAAACT 900
CATCCCTTTT TATGGCTGCA TAGTGTCCA TGGTGATAC GTGCCACATT TTCTTAATCT 960
45 ATCATTGATG GACAAGTTTT GCTATTGTGA ATAGTGCCAC AATAAACATA CGTGTGCGTG 1020
TGTCTTTATA GCAGCATGAT TTATAATCCT TTGGGTATAT ACCCAGTAAT GGGATCACTG 1080
50 AGTCAAATGG TATTTCTCGT TCTAGATCCG TAAGGAATTG CCACACTGTC TTCCACAATG 1140
TTTGAACATA TATACACTCC CACCAACAGT GTAAAAGTGT TTCTATTTT CCACAACCTC 1200
TCCAACATCT GTTATTTCCT GACTTTTTAA TGAACGTCAT TCTAACTGGC GTGAGATGGT 1260
55 ATCTCATTGT GGTTT 1275

60

(2) INFORMATION FOR SEQ ID NO: 130:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 472 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

10 CNGAAACCCC GTGAACCTC CCCGGTTAA AAAGCCCCC CTAAATGGG GGAACGCYTC 60
ACACGTTATA AAAAAGCACT AGAATGTTT GAAAGCGAGA AACAACAGCT GTGTAGGGTA 120
15 GCTAGCAGTT AGTGTGTAC AGAAGACAGA TATTTGTGCA TTTGTGCATT TTCTAAGTTT 180
GCTGCAATGA GCATGTATTA CTTTCATAGT TATAAACAC ATGCAAAATG CCCTTTTAAA 240
ATGAAAAAAA ATCCATGAGT GTAAGTGATA TATATGCTTT GAAAGCCTG GGACGGTCAT 300
20 TGTTTACTCT CAATAGTATG TGTTCGCTT TGCTTTTGT AGACATTTTG TTITAATCTG 360
TTGATGACAA TAACCTGTG ATAATATAAC TTGATAACAA ATAAATGAC TTATGATTGA 420
25 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA NN 472

30 (2) INFORMATION FOR SEQ ID NO: 131:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1950 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

40 ACCTCTCAGA ATCTTCTCTC AGCAACCTGA GTCTTGGCCG TTCTTCAGAG CGCCTCAGTG 60
ACACCCCTGG ATCCTTCCAG TCACCTTCCC TGGAAATTCT GCTGTCCAGC TGCTCCCTGT 120
GCCGTGCTG TNATTGCTG GTGTATGATG AGGAAATCAT GGCTGGCTGG GCACCTGATG 180
45 ACTCTAACCT CAACACAACC TGCCCTTCTT GCGCCTGCCC CTTTNTGCCC CTGCTCAGTG 240
TCCAGACCNT TGATTCCCGG CCAGTGTC CCAGCCCCAA ATCTGCTGGT GCCAGTGGCA 300
50 GCAAAGATGC TCTGTCCCT GGTGGTCCTG GCCCTGTGCT CAGTGACCGA AGCTCTGCCT 360
TGCTCTGGAT GAGCCCCAGC TCTGCAACGG GCACATGGGG GGAGCCTCCC GGCGGGTTGA 420
GAGTGGGGCA TGGGCATACC TGAGCCCCCT GGTGCTGCGT AAGGAGCTGG AGTCGCTGGT 480
55 AGAGAACGAG GGCACTGAGG TGCTGGCGTT GCCTGAACTG CCCTCTGCCC ACCCATCAT 540
CTTCTGGAAC CTTTGTGGT ATTTCCAACG GCTACGNTG CCCAGTATTC TACCAGGCCT 600
60 GTGCTGGCC TCTGTGATG GGCCTTCGMA CTCCAGGCC CCATCTCCTT GGCTAACCCC 660

	TGATCCAGCC TCTGPTCAGG TACGGCTGCT GTGGGATGTA CTGACCCCTG ACCCCAATAG	720
	CTGCCCCACT CTCTATGTGC TCTGGAGGGT CCACAGCCAG ATCCCCCAGC GGGTGGTATG	780
5	GCCAGGCCCT GTACCTGCAT CCCTTAGTTT GGCAGTGTG GAGTCAGTGC TGGCCCATGT	840
	TGGACTCAAT GAAGTGCACA AGGCTGTGGG GCTCCTGCTG GAAACTCTAG GGCCCCACC	900
10	CACTGGCCTG CACCTGCAGA GGGGAATCTA CCGTGAGATA TTATTCCTGA CAATGGCTGC	960
	TCTGGGCAAG GACCACGTGG ACATAGTGGC CTTGATAAG AAGTACAAGT CTGCCCTTAA	1020
	CAAGCTGGCC AGCAGCATGG GCAAGGAGGA GCTGAGGCAC CGGCGGGCGC AGATGCCAC	1080
15	TCCCAAGGCC ATTGACTGCC GAAAATGTTT TGGAGCACCT CCAGAATGCT AGAGACCTTA	1140
	AGCTTCCCTC TCCAGCTAG GGTGGGAAG TGAGGAAGAA GGGATTCTAG AGTTAAACTG	1200
20	CTTCCCTGTT GCCTTCATGG AGTTGGGAAC AGGCTGGGAA GGATGCCCAG TCAAAGGCTC	1260
	CAAGCGAGGA CAACAGGAAG AGGGATCCAC TGTTACCAA AGTCCTGATT CCCCCATCAC	1320
	CAACCTACCC AGTTTGTTCG TGCTGATGTT GGGGGAGATC TGGGGGAGT TGGTACAGCT	1380
25	CTGTCTCTCC CTGTCTCTAT ACCGGGAACCT CCCCTCCAGG GTACCCACAG ATCTGCATTG	1440
	CCCTGGTCAT TTTAGAAGTT TTTGTTTAA AAAACAACCTG GAAAGATGCA GAGCTACTGA	1500
30	GCCTTTGCCC TGAATGGGAG GTAGGGATGT CATTCTCCAC CAATAATGGT CCCTCTTCCC	1560
	TGACGTGCT GAAGGAGCCC AAGGCTCTCC ATGCCTTTCT ACCTAAGTGT TTGTATTTTA	1620
	TTTTAAATTA TTTATTCTGG AGCCACAGCC CCCTTGCTTA TGAGGTCTT ATGGAGAGTG	1680
35	AGAAAGGGAA GGGAAATAGG GCACCATGGT CCGGTGGTTT GTAGTTCCTT CAAAGTCAGG	1740
	CACTGGGAGC TAGAGGAGTC TCAAGCTCCC CTTAGGAAGA ACTGGTGCCC CCTCCAGTCC	1800
40	TAATTTTCT TGCCTGCCCC GCCTTGGGA ATGCCTCACC CACCCAGGTC CTGACCTGTG	1860
	CAATAAGGAT TGTTCCTGC GAAGTTTGT TGGATGTAAA TATAGTAAAA GCTGCTTCTG	1920
45	TCTTTTTCAA AAAAAAAAAA AAAAAAACT	1950

(2) INFORMATION FOR SEQ ID NO: 132:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 990 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

TGGAAGATTT AAAATAGGTT TCATATTTCT CTTGAATATG AATATATAAG CTTGAATAAG

60

	CTTGAGTCCCT TATTATTATG AATTTTCCT TATTATTTCT ACCAATGCTT CTTATATTAA	120
	AGCCTGATCT TTTCATATT AGTATATGTA CATTAGCTGC CTGTGGATTA ACATTTCCAT	180
5	GAATGTAAT TTTCATTTT TCGATCTTAA ACTTTTGTG TCTTTATATA AGGTATGCTY	240
	CTTTAAGCA TGATATTTT AACACAATA GTTGAAAGAC AATCTYCACC TTTTACTTGT	300
10	ATATTACAT GTATGGTAAT TTTTGATGCA TATTACGTCT TATTATTTAA CCAACCTATT	360
	TTCATTATC TAGGGCATTT TCCAGAAAGC CTTATTTTCT TGTATTATC AAATATTTT	420
	AYCATGTAT TTTCCTAT TATTTAGKAA TACGKTACYC YAAATATATA TTGTGGSTAT	480
15	TTTCAGAAAT GCATATGCC TCCTTAATTT ATTAGAGGCT AACCTAAAT ATTACTTTTA	540
	CCACTTACTT GAAATCTTG GATCTTAGA ACATTTATTG TTTTATGCAT TTTAATTTCTA	600
20	CTTGATTTT TACTACTCTT AACATTATT ATGTGTTTAG ACAAGCCAAA ATATATNTTG	660
	TTATATCTT ATCTCTAT TCCTCTGTA TTTTATGCC ACTATGTATG CTCATTTTCC	720
	TTCTATGGA TGACCTAAT TCGTACTTT TGTTTTTTAA TCTGTGCAGG TAGCCTGGCC	780
25	ATTAATTTT TATTTTGGT TTGCTGAAA AATGTGTTT ATTCTATAT GCATACTTAT	840
	GCAATAGAA TCTAGTTTG ACATATTTT AGTATTTATA AATGTAAAGT CATTWATTKG	900
30	GCTCTATCA TTCTGATGA GAATCAATT GTCAGCCCAA TAGTTTTTCA TTTTAAATTA	960
	CNGAATTTT TCACTCTCT GGTTTTAGGA	990

35

(2) INFORMATION FOR SEQ ID NO: 133:

(i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 1720 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

45

	GTCTGACAAG CGACTGTGGT TATCCCTTA AGTTTACTT CAGCACTAAC ACTAGTGCTT	60
	CCGCTGGAGT TTGCATTTT CCAGCTTTAT ACAGGATTTT CCTTTGACTG GAAGAGTCAA	120
50	GGATATGAG ACTCAAGT GATTTTATT GTACAACATC AAGGGGAATA GGATACTCAT	180
	CAACTGGGA TTATCTTAT CAAACATGG TCTTCTTGA ATAAGAAAAA TACATAGTTG	240
55	GTTATTATGG ACTATAACT GGTATAATG GATATCTGA TAAATATTT GCTGCTCTGT	300
	AGTGTGGA AAATCTGAG ATATTAGCTT TACTCATCTT GAGCTTTGAG GATGTTCTCT	360
	GTCGCCGAT GGTTCATAT TACTAAAAA AGCTGGGTAT TGTAAATCT CATTTATAAA	420
60	AATCAGATG AGAAGAAAT TCTCTTGAT GGTGAGACTG TTGCTTAGT TCAGGAAAT	480

385

ATTTAATAAT CCTTTGTTAC CTGTGAATGA AGGAACTTTG TAATTCTGAT TTATCGTAAA 540
ACATGAGCCT TTCCAGAGTC AGCTTAGACA CTGTTGTCGC AAATAGCCAT GCTTTGCCTT 600
5 ATGCCAAGGA GCGCCAGAGG GAGGGCCTAG TCTTCCTCTG TTGCTGTACA TATATTGAAA 660
TGCTTTTTTT TTTTATTTTG CATTTGTTAT CTATAATGAG CTTTCTGAGC CCTGATATTA 720
10 TGTGAGACAA ACAGGAGTTA TTGATGTTAT ACACTCCCTT CCATTCAGGA TTTTCTGCTT 780
GGAGGGAAAT ATGTTGACCT TAGAGAATTG TGAATATTGT TGCAATTCTT GAATATATTA 840
CCATGTGAAT AATAGAGACT GTGTTGCTCT CTAGTATAAG CTATATTTAT TTTTGATTCA 900
15 TTTGAATTAC TAGTTATAAC TGGAGAAAT TTGTTACCTC TATCCTGGCT TGCCTGACTG 960
GCTGTATAAT AGCAGCAGCC TCTTTTAGAG CATCTTAATG AAAACATGGA TGAAAGGAAT 1020
20 TAATGATGAT ATCTGCAGAC TGGCTAGAAA ATGGCTTTTG TTCCAGCGT TAACATTTTC 1080
TTCTCAATCA CATTTCAATG TTTGTGGAGA GTGGCAGATT CACACCAGAA ACACTAGGTG 1140
TTCATATCCA TAGCATGGAT GCAGAATAAG CAGTTGGGAG AGAAGCTTCT TCCTACCTGG 1200
25 TACTCCTCCC ATTCACCTCA GCCCAGCCCC AGACAGGCGT TAGCATTGAG TGTGGGCCCT 1260
CAGGCAGCCC TGAAGCCTGG CTGGGTCATC AGATGGGGGC AGCCTGTGAC GGCACCCAGC 1320
30 GGCTGATTC CAGGGAAGAG TTCCTGGAGG GTGTTGGCTG TTTTGTGTTAG CTCAGTTTTT 1380
TTCTGGGCTC CACCATTCTT AACTCCAGGT AGACAAGATA GATGTCACAC ACAACAATTT 1440
TAAAGTATTT TGCTTAGTGC ATTTTGTTTA TGATTGCAGT GTTTGTTTCT TATTTAATAG 1500
35 GCTTTTTACT TCATTCTATT AAATTTTAGT GTTTAGAAGA GGCGGGTACT GTCAGTGTGT 1560
AAAATATGTA ATATTTTATA TGTTATACCA TGTCATATAT ACTTGCAATA TCAGACCTTG 1620
40 CATTCATAT ACAATGCAAT TGACTCTTG CAGACCTGCA TTTTTCAGTG AACAATAAAA 1680
AGATTGTCTG GCACTCCAAA AAAAAAAAAA AAAAAAAAAA 1720

45

(2) INFORMATION FOR SEQ ID NO: 134:

50

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 705 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

GGCAGGAGC CATCTGGGCT CATTCAGCAG GAAATAATGG AAAAGCTGC AATATCCAGG 60
TGTTTACTAC AATCTGGAGG CAAGATCTTT CCTCAGTATG TGCTGATGTT TGGGTTGCTT 120
60

386

GTGGAATCAC AGACACTCCT AGAGGAGAAT GCTGTTCAAG GAACAGAACG TACTCTTGGA 180
TTAAATATAG CACCTTTTAT TAACCAGTTT CAGGTACCTA TACGTGTATT TTTGGACCTA 240
5 TCCTCAITGC CCTGTATACC TTAAAGCAAG CCAGTGGAAC TCTTAAGACT AGATTTAATG 300
ACTCCGTATT TGAACACCTC TAACAGAGAA GTAAAGGTAT ACGTTTGTNA AATCTGGGAA 360
GACTTGACTG CTATTCCATT TTGGGTATCA TATGTACCTT GATGAAGANG ATTAGGTTGG 420
10 GATACTTCAA GTGAAGCCTC CCACTGGAAA CAAGCTGCAG TTGTTTGTAGA TAATCCCATC 480
CAGGTTGAAA TGGGAGAGGA ACTTGTAATC AGCATTCAGC ATCACAAAAG CAATGTCAGC 540
15 ATCACAGTAA AGCAATGAAG AGCAGTTTTC CAATGAAAAC TGTGTAAATA GAGCATCAAC 600
AAGTACAAAA TTCTTGTCTT AATTAGTGGG GGTATATAAA AATTCCTTGT AATGGTCAAA 660
TATTTTTTAA AATTGACATT AATAAAGCAT ATTTTAAAAG TTTCT 705
20

(2) INFORMATION FOR SEQ ID NO: 135:

25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 323 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

30

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

AGCACACACC TCCTTTAGTT GCTCCTAAGG TCATGTTCAA CATTGCTGGA GTGCATTTTC 60
35 TGCTCAGGGA GCTTTCCAG ACCCGGAATG TTTGGTGCTC ACAGACYCTG GCAAGGATCG 120
GTATTGCTGT TCCTCAGTTT TGCTGGGGA AATGGAGGST CAGTGACGTT CAGTGACGTG 180
40 CCCAGAGTCA TGCCATTGGC GGGTGGCCCA GKGMTCCAGG TCTCCAGCAC CCCTCGGCCC 240
CCTCCTCACC AGGTCACATC ATCTCCTGGA TTAGAATCTG CTCACATAGT CTGTCCTGAA 300
AGGAAAAAAA AAAAAAAAAA AAC 323
45

(2) INFORMATION FOR SEQ ID NO: 136:

50

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 582 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

55

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

GGACGGAATG GTGCAACCCT CCTWAMTTT CTKGKGCTGT TGACAACAGA GGGAGGGAGG 60
60

387

GAAAACATTT TTYGTGGGAG AATCCTACT CTGCAGSGGA GCCCTTAAGC GATKGATTTT 120
 GAATCTKGAC CCTTTACCAA CTAATTTTGA AGGAAGATAC CTGGGAAATA TTTGGCATTG 180
 5 AGTGGGTTAC TGAACAGCA TTAGTGAATT CATCTAGAGA ACTCTTTCAT TTATTCAGGC 240
 AACAACTGTA CAACTTGGAA ACCTTGTTAC AGTCCAGTTG TGATTTTGGG AARGTATCAA 300
 CTCTACACTG CAAAGCAGAC AATATTAGGC AGCAGTGTGT ACTATTTCTC CATTATGTTA 360
 10 AAGTTTTCAT CTTCAGGTAT CTGAAAGTAC AGAATGCTGA GAGTCATGTT CCTGTCCATC 420
 CTTATGAGGC TTTGGAGGCT CAGCTTCCCT CAGTGTGAT TGATGAGCTT CATGGATTAC 480
 15 TCTGTATAT TGGACACCTA TCTGAACTTC CCAGTGTAA TATAGGAGCA TTTGTAAATC 540
 AAAACCAGAT TAAGGTTTGA CTGGTTTCAT TTGATTTTAA AG 582

20

(2) INFORMATION FOR SEQ ID NO: 137:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1021 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

TTCGGCAGAG CCCTTGGCG CTCTGAATA CCTGCKTTCT GTAGCGCTAG TTCTCTTCAA 60
 GATTGCTTA GTGTCATTTT ATTTGGTIT CTITTCCTGC CATGTTTTTC TGTCGGAATT 120
 35 ACGGTTCTGT TTGGTCTAT GTACTCTCTA AAATGTTATC GTTTTTTATT TGTCTACTAA 180
 TTTTCGTGCA TTGTTACTA CTGAGTTTCT TAATATCTGA CTGGCCTCCG CCCACGGGCT 240
 40 CTGCAGANCA TAAATACTC AGGCTGATGG TAGTGCAGAG ACTCTCCCTC CTTGATCAGC 300
 GCAAACGTTG GTCTGAGGCT TGAGGGATGG AGCAACATTT TCTTGGCTGT GTGAAGCGGG 360
 CTTGGGATTG CGCAGAGGTG GCGCCAGAGC CCCAGCCTCC ACCTATTGTG AGTTCAGAAG 420
 45 ATCGTGGGCC GTGGCTCTT CCTTTGTATC CAGTACTAGG AGAGTACTCA CTGGACAGCT 480
 GTGATTTGGG ACTGCTTTCC AGCCCTTGCT GGCGGCTGCC CGGAGTCTAC TGGCAAAACG 540
 50 GACTCTCTCC TGGAGTCCAG AGCACCTTGG AACCAAGTAC AGCGAAGCCC ACTGAGTTCA 600
 GTTGGCCGGG GACACAGAAG CAGCAAGARG CACCCGTAGA AKARGTGGGG CAGGCAGARG 660
 AACCCGACAG ACTCAGGCTC CRGCAGCTTC CCTGGAGCAG TCCTCTCCAT CCYTGGGACA 720
 55 GACAGCAGGA CACGAGGTC TGTGACAGCG GGTGCCTTT GGAACGCCGC CATCCTCCTG 780
 CCTCCAGCC GTGGCGCCAC CTCCCGGTT TCTCAGACTG CCTGGAGTGG ATTCTTGGCG 840
 60 TTGGTTTTGC CGCGTCTCT GTACTCTGGG CGTCTGTTC ACGGATCTGT GGAGCTAAGC 900

	AGCCTTAGAT AGCAGCAGAA GGCTTTTGG ATTCTCCTCC TTGAAAAGAT TCTCAGTTAC	960
	CAAACGTCTC CACCTAGAAA ATAAAAATAC ATTAAGATGT TGANAAAAAA AAANAAAAAA	1020
5	A	1021
10	(2) INFORMATION FOR SEQ ID NO: 138:	
	(i) SEQUENCE CHARACTERISTICS:	
15	(A) LENGTH: 1777 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:	
	CGGAAGATGA TGGCTTCAAC AGATCCATTC ATGAAGTGAT ACTAAAAAAT ATTACTTGGT	60
	ATTCAGAACG AGTTTAACT GAAATCTCCT TGGGGAGTCT CCTGATCCTG GTGGTAATAA	120
25	GAACCATTC AATACAATG ACTAGGACAC GAGACAAGTA CCTTCACACA AATTGTTTGG	180
	CAGCTTTAGC AAATATGTCG GCACAGTTTC GTTCTCTCCA TCAGTATGCT GCCCAGAGGA	240
30	TCATCAGTTT ATTTTCTTTG CTGTCTAAAA AACACAACAA AGTTCTGGAA CAAGCCACAC	300
	AGTCCTTGAG AGGTTGCTG AGTTCTAATG ATGTTCTCTT ACCAGATTAT GCACAAGACC	360
	TAAATGTCAT TGAAGAAGTG ATTGGAATGA TGTTAGAGAT CATCAACTCC TGCCTGACAA	420
35	ATTCCTTCA CCACAACCCA AACTTGGTAT ACGCCCTGCT TTACAAACGC GATCTCTTTG	480
	AACAATTTG AACTCATCCT TCACTTCAGG ATATAATGCA AAATATTGAT CTGGTGATCT	540
40	CCTTCTTTAG CTCAAGGTTG CTGCAAGCTG GGAGCTGAGC TGTCAGTGGA ACGGGTCTG	600
	GAAATCAITTA AGCAAGGCGT CGTTGCGCTG CCCAAAGACA GACTGAAGAA ATTTCCAGAA	660
	TTGAAATTCA AATATGTGGA AGAGGAGCAG CCGAGGAGT TTTTATCCC CTATGTCTGG	720
45	TCTCTGTCT ACAACTCAGC AGTCGGCCTG TACTGGAATC CACAGGACAT CCAGCTGTTT	780
	ACCATGGATT CCGACTGAGG GCAGGATGCT CTCCCACCCG GACCCCTCCA GCCAAGCAGC	840
	CCTTCAAGTT CTTTATTTT TGGGTAACAG AAGTAGACAG ACAGGTTACT TGGTGATCT	900
50	TCTGTTAAAG AGGATTGCAC GAGTGTGTTT TCCTCACACA CTTTGATTG GAGAAITGGT	960
	GCTAGTTGGC AATAGATAAC TCAGCGTAGA TAGTATTGCA AAAAGGGGAG GAAATACACA	1020
55	ACAATAATAA ATGTAAAAAC CTGCTATTCA ACATGCAGTT TTATTTCGAR GCCAAAAATC	1080
	TAGAGCTTTC CCAAGATCCT GTGCTTAG GCACATNCAC ACTTCAACAG TGCACACTAT	1140
60	CCAACAGTGC AACTATTCA ACAGTCACA CTATTCAAAA GCGTAGACTA TTTTTTTGCA	1200

389

TGTTCAGAT ATTGTTTTG GTCTTATGTG TGTGTGAGAG AGAGAGATTC CTTTGACATT 1260
 AAGGAGCATC AATGAGAAAA GATGATGAGG CAGGAATTAA TAAAGAAATG AAGTCGTGTG 1320
 5 TGTTTGGTTG CCTGTCAGAG GGCACACAAT TTCATAAACA CCATGCCCTGG ACAATTTGAT 1380
 ATTAATATTT AACACCTCTG CATCTTTTTC TTAAAAAAGA ATATGGGCCA GATACAGTGG 1440
 CTCACATTTG TAATCCAGC ACTTTGGGGA GCCAAGTTAG CAGAATCCCT TGAGCACAGG 1500
 10 AATCTGAAAC CAGCTTGGGC AACATAGTGA GATCCCATCT NTACAAAAAA CTAAAAAAT 1560
 AGCCAGGCAT GATGCCACAT TCCTGTAGTC CTAGCTACTC AGGAGGCTAA GGTAGGAGGA 1620
 15 TTGCCTGAGC CCAGGAGTTC AAGGCTGCAG TGAGCTAAGN ACGTGCCAGT AACTCCAGC 1680
 CTGAGCCACA AAGTGAGACC CTGTCTCGCA AAAAAAAN TTAAAAAGTC GGGGGGGGGC 1740
 CCGGTACCCA AATCGCCGGA TATGATCGTA AACAATC 1777
 20

(2) INFORMATION FOR SEQ ID NO: 139:

25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 643 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

30

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTGGG AATGAGAAAA TAACTTTATT 60
 35 TTCAATTGTG GGAGCGGGCC GATGTCCAGC CTCAGAACTT CTGGAAGTGC TTCTTGGTGC 120
 CGGCAGCCTT GGTGACCTTG AGCACGTGTA AGCGCACTGT CTGCTCAGA GGCGGCACT 180
 40 CGCCCACTGT GACGATGTCA CCGATCTGGA CGTCCCTGAA GCAGGGGGAC AGGTGTACAG 240
 ACATGTTCTT GTGGCGCTTC TCGAAGCGGT TGTACTTGCG GATGTAGTGC AGATAGTCTC 300
 GGCGGATGAC AATGGTCTTC TGCATCTTCA TCTTGGGTCA CCACGCCAGA GAGGATCCGC 360
 45 CCTCGAATGG ACACATTACC AGTGAAGGGG CATTTCTTGT CAATGTAGGT GCGGCTCAAT 420
 AGCCTCCTTG GGTGTCTTT GAAGCCAGA CCGATGTTCT TGTAGTAAC CCGCGGAGC 480
 50 TTCTCCTTGC CAGTTTCTCC CAGCAGGACC CTCTTCTTGT TTTGAAAGAT GGTCCGCTGC 540
 TTTTGGTAGG CAGCTCAGT CTGAATGTCC GCCATCTTCT CGTGCCGMAV TCCTGCAGCC 600
 CGGGGATCC ACTAGTTCTA GAGCGGCCG ACCGCGGTGG AGC 643
 55

(2) INFORMATION FOR SEQ ID NO: 140:

60

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1220 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

GGCACGAGGA TGATAGACCT ACTGGAGGAA TACATGGTTT ACAGGAAGCA TACCTACATR 60
AGGCTTGATG GCTCATCCAA GATCTCGGAG AGGCGAGACA TGGTTGCTGA TTTTCAGAAC 120
AGGAATGACA TCTTTGTGTT CCTGTTAAGC ACACGAGCTG GAGGACTGGG TATCAATCTC 180
ACTGCTGMAG ACACAGTGCA TTTTCTATGA TAGCGACTGG AACCCCACTG TGGACCAGCA 240
GGCCATGGAC AGGGCCCAACC GCTTAGGGCA GACAAAGCAG GTTACTGTGT ACCGGCTCAT 300
CTGTAAAGGC ACCATTGAAG AACGCATTCT GCAAAGAGCC AAGGAGAAGA GTGAGATTCA 360
GCGGATGGTG ATTTCAAGTG GGAACCTCAA ACCAGATACC TTGAAACCCA AAGAGGTGGT 420
TAGTCTTCTT CTAGACGACG AAGAGTTGGA GAAGAAACGT ATGTACTCTA AACCTCTATA 480
CACTCCCCCTC ACGTATCTGA GAATGGAAGA GGTACTTGGG TGTGTGCCAA GGGTTAGGCA 540
AAGCCAGAGG CTGTATTTAG GGAAAGTATT TTTGTGCTCA TATTTTATAT AAAAACCCAA 600
ACAAGAATGT GTTGTAGGC CAGGCGTGGT GGCTCGGCC TCTAGTCTCA GCATTTCGGG 660
ARGCCAAAGT GGCAGATCA CCTGARGTCA GGARTTTGAG TTTGARACCA GCCTGGCCMA 720
CGTTGTGAAA CCCCACCTCT ACTARGARTA CSGAAAATTG GTTGGGCATG GTGGCGGGCA 780
CCTGTAATTC CAGCACTTTG GGAGGCTGGG GCAGAANAAT TGCTTGAGCC CAGGAGGTGG 840
AGATTGCGGT GAGCCGAGAT YGTGCCATTG CAMTCCAGCC SGGGCAATAA GAGTGAAAYT 900
CCATCTTTTA AAAACAAACA AAAACAAAAA ACACAAGACG GCTCACACCT GTAATCCCAG 960
CACTTTGGGA RGCCGARGCA GGTGGATCAC GARGTCAGGA GTTCCAAGAC TAGCCTGGCC 1020
AACCTGGTGA AGCCCCGTCT CTAATAAAAA TACMAATATT AGTGGGCGT GGTGGTGGGC 1080
ACGTGTAATC CCAGCTACTC GGGAGGCTGA GGCAGGAGAA TCCCTTGAAG CTAGGAGGCA 1140
GAGGTTGCAG TGAGCCAGGA TCGTGCCAAT GCACTCCAGC CTGGACAACA AGAGCAAGAT 1200
TCCATCTCAA AAAAAAAAAA 1220

(2) INFORMATION FOR SEQ ID NO: 141:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 721 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

5 AATTCGGCAC GAGCCAGGTT AGCCGGAAGG GCAGCTCTCC AGGCCCTGCC CACCCACAG 60
 GGGGCTCCTT ATGCACAGCG GGGCGTCTCC TTGTGGCCAT AGAAACGGAA CTGGCTCTTT 120
 TCAACAGTGC TGCAAGAGGA TGGTTATTTA ACGCTGGCCC CCAAGGAGGA AAGGCACAGA 180
 10 CYTTCCTCCC TCCTGGAACA TCCAAGGGCA CTGGATCCTC TGTGTCCCTC TGAGATGGGG 240
 TGCCACTCCA GCAAGAGCAC CACGGTGGCA GCTGAGTCCC AGAAGCTTGA AGAAGAGYGC 300
 GAGGGAAGAG AGCCAGGTCT GGAGACCGGC ACCCAGGCAG CAGACTGCAA GGATGCCCCG 360
 15 CTGAAGGATG GAACCCCTGA GCCAAAGAGC TGAAATGCCT CTCTCCAGAG TCGGACCCCTC 420
 ACCTCYTTCC TGGAACTGCC TTTGGCCCCA GAACCATGAG ACAATCCCCA CCCTGAGAAG 480
 20 CTCCGATCAC TGGGAGGAGA GAGAAAGCCT CCAGCTTTGG GATTCAGGCT TCAGAAGTTT 540
 TTAGCAGCCT TTGCTCATTG GAGAGGTGGG GAAAGGATAA AGTTCTTATA AGGAAATCCC 600
 TAATTTCCCC CAGCTCCTCC CCNCCNGAAG AAGGAACNAA AGAAAGTTCC TTCCACACGT 660
 25 TTTGTGGGAA ACTTTTCCCT TGCCAACCTT CTTGGATTG CCAGAACAAA GCCCTCCAGA 720
 A 721
 30

(2) INFORMATION FOR SEQ ID NO: 142:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1468 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 40 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

ATGAATTAAT GTTTATAAAT GACTGTACTG AATTTAAAC CGTACAGTTT CATTTGCATT 60
 45 TTGACATTAC TTTATTATAC ATTTTGCATT TAAAGGCTG CACCAGTTGG CTTTCTTCT 120
 GTTTATCTCT CAAATATAG AGATTCTGTG ATTTATTTGC CCTGTTTATG GATTAAAAAG 180
 AAAATTCTAA TATAAGCAT TTCAATAGGA TGCATAGGTA TATTAGTTT TTTAAATGCT 240
 50 TTAGATCTGT GATTCTTGAC TTACTATTTA TTTTATCCCC TTAAAGTCAG GGATGCTTTA 300
 TTCTATTTTA AAGCACTTAT GAGTTACATG TTGTAATCAA GTTGCACAA TATATTTATC 360
 55 TATATGAGGA ACCATAAAT GAATAGCTAA TTTTAAAAAT GCCATTAAAA TGCATGAAAT 420
 KCTTATTAAA ACCTTACTAT ACTATTTCTT CAAGGCAAGT AAATTGACCA TGRGAAAGR 480
 ACACAGTTAT TAAACACTGT TGACAGGAAA ATTCTCCTTG ATAACATAGG ACAATTAATG 540
 60

392

GAAAAAAAA TTCTCATTAT TTGCAAAGAA TGAACAAGTT AATGAACAAA CAAACTAGAT 600
 TTGGTATGTT TTCAGCTTTT GTATCATGTT TAATTGTTTA ATTTGGTTGA AAAACTGCAG 660
 5 TTGAGAAATC AGATAGCAAT ATAGACATT CACAGCAGCTC TGTGGATACC ATGTAATTGT 720
 CAGGTAATTT CAGAAATGTTG AAAATTATT CAGTGCAGCCC TCATAGTATC ATACTTGAAG 780
 AAATTGATTA CAGTCCACT AAATTGTTGA AGATAAATTA TTTTAAAGG TTATGAAAAC 840
 10 TAAGTTATAT TAATTCATAT GTTTGATTTT TAAATCCCAC CTCTCAAGC TATCCAATTT 900
 NCTGACTTTG AAAATAACCA TGAGAGATGC CACATTTCTC TCTGGGAAAC TACCACTCAA 960
 15 AGAATAATTG TTAAAAATTA AGCTTTTAGG TATTAGAAGC TGTATAAAG TATAAAATTA 1020
 AGATATAAGC AGATCACATG TAAATCATTC CTAAAGCACA AGAAAAGAAT GTGCCTTGAT 1080
 GTACATATAT TACTAAGTTG CCTCTCCAG TTTACTTTAA AAATGGCTTT AAGGATAAAG 1140
 20 AATAAATGTTG ATAGCTGTGC ATGCATTATA TATTGCATT TGCAAATTC CCATTGTTTT 1200
 AACAGCTGTG TGGCTGACTT TCAATTTTAA GACGTGAAT GACATACAGC CCATAACTTT 1260
 25 ATAATGGCTG CTCAATTTATC TTATCTTTCA GTTAGTGGAA AAACATTTCA ACCTGACTAA 1320
 AATTGGAAT TGTGCTTTT ATGTTCCATC CTCTGTTGTT ACTAGATTTA GTTTAAAAAT 1380
 TGTGTATGAC CATTAAATGA TGCATAAAC ATGTAAATAA AAGATGTTGA ATCTTGTTGA 1440
 30 AAAGCAWRAA AAAAAAAAAA AAACCTCGA 1468

35

(2) INFORMATION FOR SEQ ID NO: 143:

(i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 300 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

45

TGAATTTTTT GCCAACTTA GAACTCTGT TAAATATTG GAGGATTTAA AGAACATCCC 60
 AGTTTGAATT CATTTCAAAC TTTTAAAT TTTTGTACT ATGTTGGTT TTATTTCTCT 120
 50 TCTGTAATC TTTGTATTC RCTTATGCTC TCGTACATTG AGTACTTTTA TTCCAAAAC 180
 AGTGGGTTTT CTCTACTGGA AATTTTCAAT AAACCTGTCA TTATGCTTA CTTTGATTAA 240
 AAAAAAAAAA AAAAAAAAAA AAACCCNAG GGGGGGGCCG GGTNCCCAAT CCCCCCAA 300
 55

60

(2) INFORMATION FOR SEQ ID NO: 144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2243 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

10 TGCCTCCCTT CCTGCAGATT GTGGACAGTA GTTCTCAGC CTGCACCCTG GATTCCTTCT 60
 TCCCCTTCCT AGCTCCATGG GACTCGCCCC AAGACTGTGG CTTCAAGGAC CACCAGCCCC 120
 TTACTCTTCA AGCCCTGACT GTGGAGTTGG TAGATGCCTC TGATCCTCAG TATTCTCTCT 180
 15 GGCAATGPTC CACGGCTTCT CCTTCCTGGG AGCTGGCTCC ATAACCTGAT TTCCCCCAA 240
 CGTGTGCAA TCCCTGCTGC CCTTAGCCA CCCAGGTCT TGTGTGGGTA TGAGTGTAGA 300
 20 GGATGGGGGT ATGCCAGGCC TGGGCCGTCC CAGGCAGGCC CGCTGGACCC TGATGCTACT 360
 CCTATCCACT GCCATGTACG GTGCCCATGC CCCATTGCTG GCACTGTGCC ATGTGGACGG 420
 CCGAGTCCCC TTYCGGCCCT CCTCAGCCGT GCTGCTGACT GAGCTGACCA AGCTACTGTT 480
 25 ATGCGCCTTC TCCCTTCTGG TAGGCTGGCA AGCATGGCCC CAGGGGCCCC CACCCTGGCG 540
 CCAGGCTGCT CCCTTCGCAC TATCAGCCCT GCTCTATGGC GCTAACAACA ACCTGGTGAT 600
 CTATCTTCAG CGTTACATGG ACCCCAGCAC CTACCAGGTG CTGAGTAATC TCAAGATTGG 660
 30 AAGCACAGCT GTGCTCTACT GCCTCTGCCT CCGGCACCGC CTCTCTGTGC GTCAGGGGTT 720
 AGCGCTGCTG CTGCTGATGG CTGCGGGAGC CTGCTATGCA GCAGGGGGCC TTCAAGTTCC 780
 35 CGGGAACACC CTTCCCAGTC CCCCTCCAGC AGCTGCTGCC AGCCCCATGC CCCTGCATAT 840
 CACTCGCTA GGCTGCTGC TCCCTATTCT GTACTGCCTC ATCTCAGGCT TGTGTCAGT 900
 GTACACAGAG CTGCTCATGA AGCGACAGNG GCTGCCCTTG GCACTTCAGA ACCTCTTCCT 960
 40 CTACACTTTT GGTGTGCTTC TGAATCTAGG TCTGCATGCT GGCGGGGCT CTGGCCCAGG 1020
 SCTCTGGAA GGTTCCTCAG GATGGGCAGC ACTGCTGGTG CTGAGCCAGG CACTAAATGG 1080
 45 ACTGCTCATG TCTGCTGTCA TGAAGCATGG CAGCAGCATC ACACGCCTCT TTGTGGTGTC 1140
 CTGCTCGCTG GTGGTCAACG CCGTGTCTTC AGCAGTCTG CTACGGCTGC AGCTCACAGC 1200
 CGCCTTCTTC CTGGCCACAT TGCTCATTTG CCTGGCCATG CGCCTGTACT ATGGCAGCCG 1260
 50 CTAGTCCCTG ACAACTTCCA CCTGATTCC GGACCTGTG GATTGGGCGC CACCACCAGA 1320
 TCCCCCTCCC AGGCTTCTCT CCTCTCCCA TCAGCAGCCC TGTAACAAGT GCCTTGTGAG 1380
 55 AAAAGCTGGA GAAGTGAGG CAGCCAGGTT ATTCTCTGGA GGTGGTGA TGAAGGGTA 1440
 CCCCTAGGAG ATGTGAAGTG TGGGTTGGT TAAGGAAATG CTTACCATCC CCCACCCCA 1500
 60 ACCAAGTTCT TCCAGACTAA AGAATTAAGG TAACATCAAT ACCTAGGCCT GAGAAATAAC 1560

394

CCCATCCTTG TTGGGCAGCT CCTGCTTTG TCCTGCATGA ACAGAGTTGA TGAAAGTGGG 1620
 GTGTGGGCAA CAAGTGGCTT TCCTTGCCTA CTTTAGTCAC CCAGCAGAGC CACTGGAGCT 1680
 5 GGCTAGTCCA GCCCAGCCAT GGTGCATGAC TCTTCCATAA GGGATCCTCA CCCTTCCACT 1740
 TTCATGCAAG AAGGCCAGT TGCCACAGAT TATACAACCA TTACCCAAAC CACTCTGACA 1800
 GTCTCTCCA GTTCCAGCAA TGCTAGAGA CATGCTCCCT GCCCTCTCCA CAGTGTCTGCT 1860
 10 CCCCACACCT AGCCTTTGTT CTGGAACCC CAGAGAGGC TGGGCTTGAC TCATCTCAGG 1920
 GAATGTAGCC CCTGGGCCCT GGCTTAAGCC GACACTCCTG ACCTCTCTGT TCACCTGAG 1980
 15 GGCTGTCTTG AAGCCCGCTA CCCACTCTGA GGCTCCTAGG AGGTACCATG CTTCCCACTC 2040
 TGGGGCCTGC CCCTGCCTAG CAGTCTCCCA GCTCCCAACA GCCTGGGGAA GCTCTGCACA 2100
 GAGTGACCTG AGACCAGGTA CAGGAAACCT GTAGCTCAAT CAGTGTCTCT WTAAGTGCAT 2160
 20 AAGCAATAAG ATCTTAATAA AGTCTTCTAG GCTGTAGGGT GGTTCCTACA ACCACAGCCA 2220
 AAAAAAAAAA AAAAAAACTC GAG 2243

25

(2) INFORMATION FOR SEQ ID NO: 145:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1082 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

GCCAAGCTCT AATACGACTC ACTATAGGGA AAGCTGGTAC GCCTGCAGKT ACCGGTTCCG 60
 40 GGAATTCCCG GGTGACCCA CGCGTCCGCT TCCGTGTGTC AAAATCCTCA CCTCCTTCAT 120
 AACCATCTCC CACAATTAAT TCTTGACTAT ATAAATTTAT GGTTTGATAA TATTATCAAT 180
 45 TTGTAATCAA TTGAGATTTC TTTAGTGCTT GCTTTTCTGT GACTCAACTG CCCAGACACC 240
 TCATTGTACT TGAAACTGG AACANCTTGG GAATGCCATG GGGTTTGATA ATCTGCCAGG 300
 GACATGAAGA GGCTCAGCTT CCTGGGACCA TGACTTTGGC TCAGCTGATC CTGNACATGG 360
 50 GAGAACAACC ACATTTTCTT TTGTGTGTGC TTCTAGCAGC TGTTGGGGAG GACCKTGACC 420
 CAAYAGTGTT CCCATGCTGT TTCTTGTGAA ATGCTCTGGG CTATGTAGCA GCTTTTGATT 480
 CCTGCATAC CCTAGGCTGC TGCCCTATC CTGTCCCTTG TTTATAACAT TGAGAGGTTT 540
 55 TCTAGGGCAC ATACTGAGTG AGAGCAGTGT TGAGAAGTCG GGGAAAATGG TGACTACTTT 600
 TAGAGCAAGG CTGGGCATCA GCACCTGTCC AGCTCTACTT GTGTGATGTT TCAGGAACTC 660
 60 AGCCCCTTTT TCTGCCTAGG ATAAGGAGCT GAAAGATTAA CTGGATCTY CTAATGGTCC 720

395

5 AAATCTTTTG GTCACAATAA AGAGTCTCCA AATTAGAGAC TGCATGTTAG TTCTGGATGG 780
 ATTTGGTGGC CTGACATGAT ACCCTGCCAG CTGTGAGGGG ACCCCGTTTT TAAGATGCAT 840
 GGCCAAGCTC TCTGCAATG GAAATGCTTA CACTGGGTGT TGGGGATGTT TGCTACCTCC 900
 TGCTATTTTT GTGGTTTTGG TTCTCCCACT ATGGTAGGAC CCCTGGCCAG CATTTGTGGCT 960
 10 TGTCATGTCA GCCCCATTGA CTACCTTCTC ATGCTCTGAG GTACTACTGC CTCTGCAGCA 1020
 CAAATTCTTA TTTCTGTCAA TAAAAGGAGA TGAAAATAAA AAANAAAAA AAAAAACTCG 1080
 15 NG 1082

(2) INFORMATION FOR SEQ ID NO: 146:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4313 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

30 CAAGCTGGTT TGAAACTAGG GGTGGGGCTC GCGCGTCGTC GTTGTTTGTC GCCGCATCCC 60
 CGCTTCCGGG TTAGGCGGTT CCTGCCCGCC CCTCTCTCTC CTCCCTTCGG ACCCATAGAT 120
 CTCAGGCTCG GCTCCCGGCC CGCCGCAGCC CACTGTTGAC CCGGCCCGTA CTGCGGCCCC 180
 35 GTGGCCACCA TGTCCCTGCA CGGCAACGG AAGGAGATCT ACAAGTATGA AGCGCCCTGG 240
 ACAGTCTACG CGATGAACTG GAGTGTGCGG CCGGATAAGC GCTTTCGCTT GGCGCTGGGC 300
 AGCTTCGTGG AGGAGTACAA CAACAAGGTT CAGCTTGTGT GTTTAGATGA GGAGAGTTCA 350
 40 GAGTTTATTT GCAGAAACAC CTTTGACCAC CCATACCCCA CCACAAAGCT CATGTGGATC 420
 CCTGACACAA AAGGCGTCTA TCCAGACCTA CTGGCAACAA GCGGTGACTA TCTCCGTGTG 480
 45 TGGAGGGTTG GTGAAACAGA GACCAGGCTG GAGTGTTTGC TAAACAATAA TAAGAACTCT 540
 GATTTCTGTG CTCCCTGAC CTCCTTIGAC TGGAAATGAGG TGGATCCTTA TCTTTTAGGT 600
 AACTCAAGCA TTGATACGAC ATGCACCATC TGGGGGCTGG AGACAGGGCA GGTGTTAGGG 660
 50 CGAGTGAATC TCGTGTCTGG CCACGTGAAG ACCCAGCTGA TCGCCCATGA CAAAGAGGTC 720
 TATGATATTG CATTTAGCCG GGCCGGGGGT GGCAGGGACA TGTTTGCCTC TGTGGGTGCT 780
 55 GATGGCTCGG TGCGGATGTT TGACCTCCGC CATCTAGAAC ACAGCACCAT CATTTACGAA 840
 GACCCACAGC ATCACCCACT GCTTCGCCTC TGCTGGAACA AGCAGGACCC TAACTACCTG 900
 60 GCCACCATGG CCATGGATGG AATGGAGGTG GTGATTCTAG ATGTCCGGGT TCCTGCACAC 950

	CTGTSGCCAG GTTAAACAAC CATCGAGCAT GTGTCAATGG CATTGCTTGG GCCCCACATT	1020
	CATCTGCCA CATCTGCACT GCAGCGGATG ACCACCAGGC TCTCATCTGG GACATCCAGC	1080
5	AAATGCCCCG AGCCATTGAG GACCCATATCC TGGCCTACAC AGCTGNAAGG WGAGATCAAC	1140
	AATGTGCACT GGGCATCAAC TCAGCCCGAA YGTGCGCCAT CTGCTACAAC AACTGCCTGG	1200
	AGATACTCAG AGTGTAGTGT TGGTGGCGCT GTGCCCACGA GGCAGGGGCT TTTGTATTTT	1260
10	CTGCCTCTGC CCCACCCCCA AAGTAAGAAG AAACATGTTT CCAGTGGCCA GTATGTCTTT	1320
	CATTGCTTTG CACCCACTGT TACCAGAAGC TGCTCTAGGA GTTCCTGGCC AGTCACCCCA	1380
15	TGCGCCTCTG TGGCAGACTC AGTGTGTGTG GCGCCTCCT CAGCCAGGG CTGAGTTTAA	1440
	AGATTTTCTC TCCTTTCTC TTCTCTTTG GTTCCTCAAT TAAAAAATGT GTGTATATTT	1500
	GTTTGTCAAG CGTGTGTGTG AGGAGCAGTT CACGCACTGG CTGTGTCTAT TCCTCTGCCC	1560
20	AGGTGTCTCT GTTGTCTGCC CAARGYWKKT TTTCATGICT CGTCCATGTC CATGTTCTGT	1620
	TTAGCACTWA CGTGGGAACA AATACCAATT TGCTTTTTCT CCTAGTATCA GTGTGTTTAA	1680
25	CAAATTTTAA CTTGTATAT TGTATATCTA TCAGGCTAAT TTTTTATGA AAAGAATTTT	1740
	ACTCTCTGCT TTCAATTTCT TGTCTATAG TCCTCCCTCT TTGCACCTTC TTCCTCTCCC	1800
	TCAGTGCTG GAGCTGGTAC TGGGCCCCCT GCCCATGAG CAGTTTGCCT TCTTGAGTCA	1860
30	CTGCCTGTGT AGTACATACC TGACCGGGAG TCCAAACCAC CTTGGTGCTC TGAAGTCCAC	1920
	TGACTCATCA CACCTTTCTT AGCCTGGCTC CTCTCAAGGG CATTCCTGGC TTGTAAACAG	1980
35	ACATAGGAAG CCTCTGTTTA CCTGAAGCA CCACTGTCCA GCCCATGGT TCCCACTGGC	2040
	AGCATGGTAG AGCTGAGAGA AACAGGCTCT CAGGGTACCT GACTTGAGGG GAATCGTTTC	2100
	ATGAAGCTGA ACTTCAAGCA TATTTCCAGT ACAATCTTTC AGAGTCTGTT TTTCCATCCA	2160
40	AATATAAGCC CCAGGCCATT CCACTTAGTG TCTTTTCAAT GATAGGCAAG AATGATATCT	2220
	GAGTTGAAC TCGGTGCTC TGTGTTTGA GTTACTGTG CCTGGTGGTA TATTGGGCAT	2280
45	TCTTTGGATT GAGTGTCTG AGGTGAGAGA GTCTTCCGA GGCATCCTGT CTGTGCTTCC	2340
	AACCTGAAC AAGACCTTAC ATGAGAGATG GACTGATGGA CTGCGGCAAT CCTGGGCTGT	2400
	CAAGTGGATA GATAGTTAAA AAGCATTATA CTGTGGGTAA TGAAAAGGA GGAAAAAAA	2460
50	AGAAGGAAAA GGAATTATAG ACCCCCAGG TCAGCCAGTT AAGAGCTCTA CCCACACCTG	2520
	TCAACCCCTC TCTCCCCAG TTTAGGTTCT GAGCAGTATT GGACTTGTAG CCTGCAGTTG	2580
55	TCTTTTGAAT TGAGGCGGC AGTGTCTTTC TGTATGTGA ATGAGTTCCA TGGAGGGGCA	2640
	TATGTGTGAT TCCACCGTTA GATGAGCCCT TGGGGCAGGC AGTTTGGGAT GTGCTCTTGG	2700
60	GGGAAAGTTG GCTGTTTCT TCGCTCTGC TCCTACCCGA AGTTTTTAAG TCCCTCTGAA	2760

397

	TTGCTCATCT GAGATTAGTA GAGTAGCAGG CCTGAAGGAT GATGGTTTTG TCCTCTTTGG	2820
	TTCTCACCTG CTGAGAAGT AAAACAGTAA CTTGTGTTCT CTGGGCCCTT AAGCTTTTTT	2880
5	GGTTAAGTCT TCCTTTTCAG AAGTAGATGT CATTATATGC CAAAAGTCTA GCTCTTTGCT	2940
	TTACCATACA GGGACCTGTC CCAAAGAAAA AGGCTCTTTT TTTAGCCAGC ATATTTCCTC	3000
10	TTCTACCCCTT TTACTTTGTT GTTCTGATTT TAGGACTCTG GCTGGCCATG TGCTTGTGGT	3060
	TGCCTCTCCT GCATTTGCCA CTGGATTTGC ACTGCATCGT TTGGAGATAC AAAGCGAGCA	3120
	GTTCTTGGTC AGAACCTTCC TCTGCTTTTC ATGTGTGTTG ATAATGGTTA CTGGGTCTCT	3180
15	CTCTCAAGGG TAGCAAGGCC AAGCTGATGG CTGCTTGTTC AGGAGGCCAT CAGTTCCTTC	3240
	CTGTGGAGAA GGGTCTGAAA TGGAACTCAG TGGTAGAAGG GGCTGGTCTG CTGGGCAGGG	3300
20	CTTACATCCA CTGAGTTCTA AGATTCTTTT CCTGATCTGC ACCTACGCCT GGTCTGTATG	3360
	GTGGAATTTG TCAGCTGGAA CTCAGAAACA ACAACTTGAA AAAAAATAA TAATTAGAAC	3420
	ATATTTGCAT AAGATAGCTA TTTACTCTGG AAACCAACAA CTTTTGAGAT TTCCCTTGCC	3480
25	CTGTGGACGC CCAGCTCCTG TCATCCTTCC TTAGGTCTCG CAGTACAGTC TTCCCTGAA	3540
	TGCCACCGGG GACCCAGGGG GACTCCACCC CCCTAAGCAA GCACACACAT ACTCACAGTT	3600
30	GATGAGTTGC TGGTCTTTGA GTCCAGCTC TCTTACCCTC CCTTTACTCC ACCAGCCCGA	3660
	CGACCCATGA CTGAGGAGGG GATTTCTACA GTCTCAGGAT TTAGAAAGTC TGTAAAGCAT	3720
	CCATGCTCCA GAAAGCACCG ATCTGTTGTA GTTGCAAAAA CAACTCTGTA ATTTGTTGAG	3780
35	GTCTCTAAAC TGACAGCCAG CGAGACTGGG TGGGAGGCC TGGATCTGTT CTCCCTGACT	3840
	GCGGGAGGAG CAGCCACTAG GACTTTAGCA GGAAGCCAC ATGGAGGCTC CGCCAGGCTG	3900
40	TGGCCAGCT GGTGATGGCC CTTTGTCTCC TGGCAGCCTG AGGCACAGCT GCCTGTATG	3960
	TCCTCATCTG TTCTGACTGA AGGATGGAGG TGCTGAATAA ATTAGGCCTC AGGCNTCTAC	4020
	CACCAGAGAG CTGGAGAATG GGTCCACGTC ATTCAAGGAC CTGAATTTTT TATGCTCAGG	4080
45	AGCATTGGAA TCCTCTTCTT CCAGGGAGGA ATTAGCCTGC AAGGTTAGGA CTTGAAGAGG	4140
	GAAGGTATTT AATAACTGGG CGAGGATGGG TGTGGTGGCT CACACCTGTA ATCCAGCAT	4200
50	TTTGGGAGGC TGAGGTGGCC AGATCCCAAG GTCAGAAGAT CGAGACCATC CTGGCTAACA	4260
	TGGTGAAACC CCATCTCTAC TAAAAATACA AAATTAAATT GGGCGGGCGT GAA	4313

55

(2) INFORMATION FOR SEQ ID NO: 147:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1183 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

5 GGCAGAGCCT CAAGCTGACT TGGATTATGT GGTCCCTCAA ATCTACCGAC ACATGCAGGA 60
GGAGTTCCGG GGCCGGTTAG AGAGGACCAA ATCTCAGGGT CCCCTGACTG TGGCTGCTTA 120
10 TCAKMYGGGG AGTGTCTACT CAGCTGCTAT GGTACAGGCC CTCACCCCTGT TGGCCTTCCC 180
ACTTCTGCTG TTGCATGCGG AGCGCATCAG CCTGTGTGTTT CTGCTTCTGT TTCTGCAGAG 240
CTTCCTTCTC CTACATCTGC TTGCTGCTGG GATACCCGTC ACCACCCCTG GTCCTTTTAC 300
15 TGTGCCATGG CAGGCAGTCT CGGCTTGGGC CCTCATGGCC ACACAGACCT TCTACTCCAC 360
AGGCCACCAG CCTGTCTTTC CAGCCATCCA TTGGCATGCA GCCTTCGTGG GATTCCCAGA 420
20 GGGTCATGGC TCCTGTACTT GGCTGCCTGC TTTGCTAGTG GGAGCCAACA CCTTTGCCTC 480
CCACCTCCTC TTTGCAGTAG GTTGCCCACT GCTCCTGCTC TGGCCTTTC TGTGTGAGAG 540
TCAAGGGCTG CGGAAGAGAC AGCAGCCCC AGGGAATGAA GCTGATGCCA GAGTCAGACC 600
25 CGAGGAGGAA GAGGAGCCAC TGATGGAGAT GGGGCTCCGG GATGCGCCTC AGCACTTCTA 660
TGCAGCACTG CTGCAGCTGG GCCTCAAGTA CCTCTTTATC CTGGGTATTC AGATTCTGGC 720
30 CTGTGCCCTG GCAGCTCCA TCCTTCGCG AGCATCTCATG GTCTGGAAAG TGTMTGCCCC 780
TAAGTTTATA TTTGAGGCTG TGGGCTTCAT TGTGAGCAGC GTGGGACTTC TCCTGGGCAT 840
AGCTTTGGTG ATGAGAGTGG ATGGTGCTGT GAGCTCCTGG TTCAGGCAGC TATTCTTGGC 900
35 CCAGCAGAGG TAGCCTAGTC TGTGATTACT GGCACCTGGC TACAGAGAGT GCTGGAGAAC 960
AGTGTAGCCT GGCTGTACA GGTACTGGAT GATCTGCAAG ACAGGCTCAG CCATACTCTT 1020
40 ACTATCATGC AGCCAGGGGC CGCTGACATC TANGACTTCA TTATTWATR ATTGAGGACC 1080
ACAGTGGAGT ATGATCCCTA ACTCCTGATT TGGATGCATC TGAGGGACAA GGGGKCGGT 1140
STCCGAAGTG GAATAAATA GCGGGCGGTG GTGACTTGCA CCT 1183
45

(2) INFORMATION FOR SEQ ID NO: 148:

50

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 734 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

55

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

60 GAATTCGGCA GAGTGAAGCA TTAGAATGAT TCCAACACTG CTCTTCTGCA CCATGAGACC 60

399

AACCCAGGGC AAGATCCCAT CCCATCACAT CAGCCTACCT CCCTCCTGGC TGCTGGCCAK 120
 GATGTGCGCA GCATTACCTT CCACTGCCTT TCTCCCTGGG AAGCAGCACA GCTGAGACTG 180
 5 GGCACCAGGC CACCTCTGTT GGGACCCACA GGAAAGAGTG TGGCAGCAAC TGCMTGGCTG 240
 ACCTTTCTAT CTCTCTAGG CTCAGGTACT GCTCCTCCAT GCCCATGGYT GGGCCGTGGG 300
 GAGAAGAAGC TCTCATACGC CTTCCTACTC CCTCTGGTTT ATAGGACTTC ACTCCCTAGC 360
 10 CAACAGGAGA GGAGGCCTCC TGGGGTTTCC CCRRGGCAGT AGGTCAAACG ACCTCATCAC 420
 AGTCTTCCTT CCTCTTCAAG CGTTTCATGT TGAACACAGC TCTCTCCRCT CCCTTGATGAT 480
 15 TTCTGAGGGT CACCACTGCC ARCCTCAGGC AACATAGAGA GCCTCCTGTT CTTTCTATGC 540
 TTGGTCTGAC TGAGCCTAAA GTTGAGAAAA TGGGTGCCAA GGCCAGTGCC AGTGTCTTGG 600
 GGCCCTTTTG GCTCTCCCTC ACTCTCTGAG GCTCCAGCTG GTCTCTGGAC ATGCAGCCAG 660
 20 GACTGTGAGT CTGGGCASGT CCAAGGCCTG CACCTTCAAG AAGTGAATA AATGTGGCCT 720
 TTGCTTCTAT TTAA 734
 25

(2) INFORMATION FOR SEQ ID NO: 149:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1405 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

GGCACAGTGG ACCCCAGACT CCCTCTCCGC CTTTCTCTGC CTGGGGAGAC CCACTGTGTG 60
 40 CATGGCATCA CTGACTCCCA TACCTCTGGC TATCAAAGGT TTCTGCCATG GCCACCCTGG 120
 AAGSAAACCA GAGGGAGGTA GACAGGGAGA TCAGGTCCCT TCTACTCTGG TTCTGTCTCT 180
 45 GTGAAATGT CTCAGGCTGG CTGTGTCCAG ARGGTCCCTG GTTCTCTCAR GGATGCCAAA 240
 TCTACAAGAA TCTCTCTCT TCCAGTTCTT ATAACCTCTC CTTCTTTTG TCTCTTTAGA 300
 CCTTGAGTA GTAGCAGCCA GGTTCCTTCT ATCTCTGGGT TAGTGCAITA TCTCTGGTGG 360
 50 CTCCCTTACC CAGGACTTTG GGAATGGTCT TTTTGTAAATA CATCTCTCTC AAATAATTCA 420
 ATTTTGAGTG TTCTGTATGT ATCTGTCTGG GAGGTGTGTA TATACAAATC ACTGTGCCCC 480
 TTTAGCAGAG AAGGAGACTG AAGCTCAGGG AGGTAAAGTG TCTTTCTCTA GGTCGTATTG 540
 55 TGGAGAAAGT GGCTGACTGG GGACTTGAAT GAGGTCCCTA GTTTCATGCT CGGAGGGCAA 600
 AGANGAATGT CCAATTGGCC TGAGATAAGC CTCTGGTAAA ATGTACTGTA CATAATAGGT 660
 60 AATCAATAAA TGTGGCTGA TGACAAACAT GTTTCTTTTG TTCATTAGTT ATAGTGATTA 720

400

TGTTCATAAT AACTCCMACA AGGAARTCAG CACATTTGGA ATATCAWTAT CTTTCCATGA 780
 TAATATCTTT CCMYGGAAAG AWAATGATAT TCCMAACTGG GAGTGTCCCW AGCARATCTG 840
 5 ANTCTGTGTA TTGGCCCTGG GGTGGGCCAG CCCCTTAGAC TCTATGGTCT CATCTCTTTT 900
 GTTACAAAA TTGAGATAAG GCCTTATTCT CTCCCCACCC CACCCATCCA TATTGTTTTG 960
 10 AGAATAAAAT GAGAGGATGT GTGTCAAGG TGTATTTTGG CAATAGTCTC TGAGCCATTT 1020
 TCTGAGCACC TCCATACTGT TGACACTCAA GTAATATTTT ATCAGCATTC CATTCAGGNT 1080
 CCTCCCTTAA TGAGGTGTGC GATGTACAAG AGTYGTGAGG TGGCAAAGGA TGGGCTCCTG 1140
 15 AGGAAACACT TAGGAAACTG GGCTTTCTGC CATTAAGA GACAAACCTT TGTGGTGACC 1200
 TAATTAAAGT TTTTAAATT CAATTTGGAA AGTTAGCAAG CTAGCTCCTK TCCAGGWAAA 1260
 20 ATAAGGAGTC AGTGCATGAC CTAACCGGC CCGGCTGCT TGCCATTCCA AACAACTGCA 1320
 GTAAGTTTAT CACNTCTTT CAGGGACTGA GGTTCAGG CACAGACTTG GATAAGGAAG 1380
 GATGTCCTAT GGGTCACAT TGATG 1405
 25

(2) INFORMATION FOR SEQ ID NO: 150:

30

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2890 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

TTATATGCTA CAGCTACAGT AATTCTTCT CCAAGCACAG AGGANTTTT CCAGGATCAG 60
 40 GGGGATCGCG CGTCACTTGA TGCTGCTGAC AGTGGTGTG GGAGCTGGAC GTCATGCTCA 120
 AGTGGCTCCC ATGATAATAT ACAGACGATC CAGCACCAGA GAAGCTGGGA GACTCTTCCA 180
 45 TTCGGGCATA CTCACTTTGA TTATTCAGG GATCCTGCAG GTTTATGGG ATCAAGCAGC 240
 CATATGGACC AAATTATGTT TTCTGATCAT AGCACAAAGT ATAACAGGCA AAATCAAAGT 300
 AGAGAGAGCC TTGAACAAGC CCACTCCCGA GCAAGCTGGG CGTCTTCCAC AGGTTACTGG 360
 50 GGAGAAGACT CAGAAGGTGA CACAGGCACA ATAAAGCGA GGGGTGGAAA GGATGTTTCC 420
 ATTGAAGCCG AAAGCAGTAG CCTAACGTCT GTGACTACGG AAGAAACCAA GCCTGTCCCC 480
 55 ATGCCTGCCC ACATAGCTGT GGCATCAAGT ACTACAAAGG GGCTCATTGC ACGAAAGGAG 540
 GGCAGGTATC GAGAGCCCCC GCCACCCCT CCCGGCTACA TTGAATTCC CATTACTGAC 600
 TTTCCAGAAG GGCACTCCCA TCCAGCCAGG AAACCGCGG ACTACAACGT GGCCCTTCAG 660
 60

	AGATCGCGGA TGGTCGCACG ATCCTCCGAC ACAGCTGGGC CTTCATCCGT ACAGCAGCCA	720
	CATGGGCATC CCACCAGCAG CAGGCCTGTG AACAAACCTC AGTGGCATAA AYCGAACGAG	780
5	TCTGACCCGC GCCTCGCCCC YTATCAGTCC CAAGGGTTTT CCACCGAGGA GGATGAAGAT	840
	GAACAAGTTT CTGCTGTTTG AGGCACAGAC TTTTCTGGAA GCAGAGCGAG CCACCTGAAA	900
	GGAGAGCACA AGAAGACGTC CTGAGCATTG GAGCCTTGGA ACTCACATTC TGAGGACGGT	960
10	GGACCAGTTT GCCTCCTTCC CTGCCTTAAA AGCAGCATGG GGSTTCTTCT CCCCTTCTTC	1020
	CTTTCCCTTT TGCATGTGAA ATACTGTGAA GAAATGCCC TGGCACTTTT CAGACTTTGT	1080
15	TGCTTGAAAT GCACAGTGCA GCAATCTTCG AGCTCCCACT GTTGCTGCCT GCCACATCAC	1140
	ACAGTATCAT TCCAAATTCC AAGATCATCA CAACAAGATG ATTCACTCTG GCTGCAC TTC	1200
	TCAATGCCTG GAAGGATTTT TTTAATCTT CTTTTAGAT TTCAATCCAG TCCTAGCACT	1260
20	TGATCTCATT GGGATAATGA GAAAAGCTAG CCATTGAACT ACTTGGGGCC TTTAACCAC	1320
	CAAGGAAGAC AAAGAAAAAC AATGAAATCC TTTGAGTACA GTGCTTGTC ACTTGTTTAC	1380
25	AATGTCTCC TTTTAAAAA AAAAAATGA GTTTAAGAT TTTGTTTACA GAGTAAATAT	1440
	ATATCCATTT AATGATTACA GTATTATTTT AAACCTTAAG TAGGGTTGCC AGCCTGGTTT	1500
	CTGAAAAACC AAATATGCG GACAGGGTGT GGCCACACCA AGAAGACGGG AAGACCTGGC	1560
30	TTGTGACCCT GGCTTCCCAT GTCTTCTGG TCTCACCOC GAAGTGCCCT ATCCTGGAAG	1620
	TATGAAATGT TAGCCAATTA ATACCAAGAC ACCTCATCTG CTCTTCCCC AGTGGATGGG	1680
35	GTCTTCTGT AAAACTGTTT GCACATGGCC AGGGGAGGGA ACTAGGACCC TTGTGTCTG	1740
	TCTGAGCCTT ATGGAGGCAG GACGGTGTCA TTGGCGGATG TGTCTGCTC CATTGAGATG	1800
	GATGGCAAAC CCCATTTTTA AGTTATATTT CTTTGATTTT TGTTAATTTA GAGGTGTAGG	1860
40	TTTTGTTTTT TGTTTTTTTG TTTTTTTTA AGAGAAACAT TTATACTGG ATAGCATTC	1920
	AGTGAAAGCA GCTGGGATG TTGGAGCTAA TGCCAGCTGT TTATACTGCT CTTTCAAGAC	1980
45	AGCCTCCCTT TATTGAATTG GCATTAGGGA ATAAACAAGC CTTTAAACGT GATAAAAGAT	2040
	CAAAAACCTG GTTAGACATG CCAGCCTTTG CAAGGCAGGT TAGTCACCAA AGACTAACCT	2100
	CCAAGTGGCT TTATGGACGC TGCATATAGA GAAGGCCTAA GTGTAGCAAC CATCTGCTCA	2160
50	CAGCTGCTAT TAACCTATA ATGACTGAAA TGACCCCTCC ACTCTATTTT TGTGTTGTTT	2220
	TGCACAGACT CCGGAAAAGT GAAGGCTGCC AATCTGAGTA GTACTCAAAT GTGAGGAAT	2280
55	GCTGGTCTTG GATTTTTTTT CCATTAAATT CAGCTGATCA TATTGATCAG TAGATAAAG	2340
	TAAATAGCTT CAAATTTTAA AAGTGAATT GCAGTGT TTTTCACTGTAT CAAACAATGT	2400
60	CAGTGCTTTA TTTAATAATT CTCTTCTGTA TCATGGCAAT TGTCTACTTG CTTATTACAT	2460

TGTC AATTAT GCATTGTGTA TTTTACATGT AATATGCATT ATTTGCCAGT TTTATTATAT 2520
AGGCTATGGA CCTCATGTGC ATATAGAAAG ACAGAAATCT AGCTCTACCA CAAGTTGCAC 2580
5 AAATGTTATC TAAGCATTAA GTAAITGTAG AACATAGGAC TGCTAATCTC AGTTCGCTCT 2640
GTGATGTCAA GTGCAGAATG TACAATTAAC TGGTGATTTC CTCATACTTT TGATACTACT 2700
TGTACCTGTA TGTCCTTTAG AAAGACATTG GTGGAGTCTG TATCCCTTTT GTATTTTAA 2760
10 TACAATAATT GTACATATTG GTTATATTTT TGTGAAGAT GGTAGAAATG TACTATGTTT 2820
ATGCTTCTAC ATCCAGTTTG TACAAGCTGG AAAATAAATA AATATAACAT AAAAAAAAAA 2880
15 AAAAAAAAAA 2890

20 (2) INFORMATION FOR SEQ ID NO: 151:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2399 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

30 GAACTTTTCC ATCTGGCAAA COGGAACTC CATCCCCATT AAACCAACTC CCCCTTTTGG 60
TTTCCCCCCC AGNGGAATAG AATTTGGACN CCCATATAAA TCCAGGAAAC CACCTAAATT 120
CTTTAGTNGT TTGTGTTTGC AAGATCTAAG GTCATGGTAA ACATTAAGTT CTTAAAATTT 180
35 TTGGGAGGGA CCGATGCACC TCTCCCTCTG AATGTTTCNC CAATTTAAAA TTGGAGTAAG 240
GTTTTAAAT GTCTNATTC ATTGGAAGGG TGTGTTATTT CATTTTGAGC CCAGAGGGGA 300
40 GAGGCACATT TTAAATATCA GAATTAGATT AGCTTTGAGT TTGTACAATT GGAACATAA 360
TAGATTTTCA TAAATTATGT GTGCCTTGTT OGAAGTGCA ACTGCTTTTA TGTCIGCTTG 420
TAAAAGTTTC AAAATATGTT TTCCCTCAAA AAGGCAACGT TACTTCATTT GCTTGAATAT 480
45 TATGATAGGA ATGCTTACTG ATATTACTTG ATAGTCATAT ATAGCCTAGG AAATTTAACA 540
TATATATAAC TATAGCAGTA TTAATAATGA TAGTTGTAAT TCTTTAAAAC ATTAAATTTG 600
50 AGGAAACTTT AATGCTGTCT CGTGACATT GCTTTACTAC AGTGAGGGGG AATATCCTTT 660
AGATTGAGCC TCAATTTACT GGTAGTAGT ATGTGAACTC TGGTATAAAA ACGTAAACTA 720
GACAGTAGAG CCGATGAATT AAAATGTGAA ATTGCTACAT TGGCATTTC TACCTCCTTT 780
55 TCTGTCAGAG TATTACTTTT TCCAGCATTT ATTCTTATTT GTGAGTAAAG AGGAAATGGG 840
AACCTGAGGT TAAAATTGAC ATTTTGTGTT CATTGAGAAT TTAAGCAGTA GGTACAGGAG 900
60 AAGTGACTTG TCACATTAAT TGGTGCCTA AATCTGTAAC TACAAGTTGT GATCGACATG 960

	TACAAAATGT CTAAGAAAGG TCATATGCTG AATATTTTAC TTTTCCTGTA TAGTCTGCAT	1020
	GATTTGTTC ATAAACCCAG CTTATTTCTT CCAAAAAGCA AAATGGTCCT GTAATTTTTA	1080
5	AAGTAAATA AACGTGCCAT TTGTCTGCA ATCTATAATT TCAGGAAGTT ATTGRAAGTT	1140
	CTGACTCAGG GCTTTTAAAC AGTTCAAGCA ATTGTCAAGT ATATTTTGGG AACTCCATCT	1200
10	GTGTAATCT CCAGTGCCCT GAAAGAATTA TTAAGTTGGC AACACTATTA AAAGTTTATA	1260
	AAAGATGGTC TTTAGTGCAC GTGTATCATT ATATACAGT TTTAAAGTCA TATGCTTAG	1320
	CTGTTAATA ATGATCTGC ATGTGTGCTG GGTGTGGTA ATTCTTTAAA GGAAGTTTTC	1380
15	TAGATTGCA CTGTATGTT GTTTTAAA AACTGATTAT TTATGGCCGT GACACTGTTA	1440
	CCAGAAAAGT AATCTAATT AAGTTATTAT GCAAAGTCAT CTATAAGTAG CATCTGGGAA	1500
20	GAGGAGATSG AGGCCACAGT TTGCTATTTT AGTATGAAAG GAGGATCTGT TTGGGAAACA	1560
	TAGATTGCT TCCCTCAAA TGAGGGGAAA AAAAAAGACC CTTGTTCAT ATGGATTCTG	1620
	TTGTAATAA TTATTTTAA AGGAAATCAC AAATGTATG TCATCTTAA TGCTAGTCTT	1680
25	ATAGAATAA TCCATAAAT TGTTTTATG TTCAGTATG TTATGTCATT CTAAATGCAG	1740
	CAAATTCAT GATAGCAGT CAATTGACTC ATAGCAGTGT TTGTATTTT TTCTAATCT	1800
30	TTAGCTTCA ATATTGGAT AAAGTCTGT TTGTGAATAT AGTTTCGTA TGGCAAATGA	1860
	TTTCTTGCTT ATTAGCTTT GTTAAAGAAT GCTTAGTAAG AGCTAAGCTT TTTAAAGTAA	1920
	TGCAACATT TATCGTAAAT AAAACCTATG GTGTAATATC ATATAATGCT TTTCTTGAT	1980
35	CTTTGGAGAA TTATCTTTT ATAGTAGTAT ACATGAATTT TGATTTTAA AGCATTTAA	2040
	AACAAATCTC AATACATTA AAAACCTGTT ATTGTAAAA RGGAAATTAC CATGCCTTTA	2100
40	AGAAACAAGG ATGTACATCT TCAATTCAGC ATAGTGTCC ACATCTAGAA GGCTCTCATT	2160
	GCAGTTGTT ACAGTTAAGG TACCTCTATC TAAAGGCCA AAGAAGCATT TCATATTTA	2220
	ACACCTCACA TTCTTTCAGG ATTAAGACAT ATGAAATAG TCTGAATAGG ATAAATTTGG	2280
45	ATAGGAAGTA ACTTAACCAG TCTGGGAAGA TTCAGGCTTT TTCTATKAAA AAGCTTATTC	2340
	CTCTTCACAA CTCNGGIGGT AGGTTTCAT TTTCAAGAG GGTAGATATT TTAAAGCCA	2399
50		

(2) INFORMATION FOR SEQ ID NO: 152:

- 55 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 802 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- 60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

CGTGCCTGTA GTAAGCTCAT CCCTGCCTTT GAGATGGTGA TGCCTGCCAA GGACAATGTT 60
 5 TACCACCTGG ACTGCTTTGC ATGTCAGCTT TGTAAATCAGA GATTNTGTGT TGGAGACAAA 120
 TTTTTCCTAA AGAATAACWT GAYCCTTTGC CARACGGACT ACGAGGAAGG TTTAATGAAA 180
 GAAGGTTATG CACCCCMGGT TCGCTGATCT ATCAACATCA CCCCATTAAG AATACAAAGC 240
 10 ACTACATTCT TTTATCTTTT TTGCTCCACA TGTACATAAG AATTGACACA GGAACCTACT 300
 GAATAGCGTA GATATAGGAA GGCAGGATGG TTATATGGAA TAAAAGGCGG ACTGCATCTG 360
 15 TATGTAGTGA AATTGCCCCA GTTCAGAGTT GAATGTTTAT TATTAAAGAA AAAAGTAATG 420
 TACATATGGC TGGATTTTTT TGCTTGCTAT TCGTTTTTGT GTCACCTGGC ATGAGATGTT 480
 TATTTTGGAC TATTGTATAT AATGTATTGT AATATTTGAA GCACAAATGT AATACAGTTT 540
 20 TATTGTGTTA CCATTTGTGT TCCATTTGCT YCTTTGTATT GTTGCATTTA GTACAATCAG 600
 TGTTTAAACT TACTGTATAT TTATGCTTTC TGTATTTACC AGCTATTTTA AATGAGCTGT 660
 25 AACTTTCTAG TAAAGAATTG AAAAGCAAAT CCTCACTAAA GGATACACAG GATAGGATAA 720
 AGCCAAGTCN CATCAACATT AAAAAATACT AAAANANAAA ACACAAAAAA AAAAAANCCC 780
 GGGGGGGGCC CGGAACCCAT TC 802
 30

(2) INFORMATION FOR SEQ ID NO: 153:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: -461 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 40 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

CTAGGAGCAC CGAGCAGCTT GGCTAAAAGT AAGGGTGTCG TGCTGATGGC CCTGTGCGCA 60
 45 CTGACCCGCG CTCTGCNCTC TCTGAACCTG GCGCCCCCGA CCGTCGCGC CCTGCCCCG 120
 AGTCTGTTC CCGCGCCCA GATGATGAAC AATGGCCTCC TCCAACAGCC CTCTGCCCTG 180
 50 ATGTTGCTCC CCTGCCGCC AGTTCTTACT TCTGTGGCCC TTAATGCCAA CTTTGTGTCC 240
 TGAAGAGTC GTACCAAGTA CACCATTACA CCAGTGAAGA TGAGGAAGTC TGGGGCCGA 300
 GACCACACAG GTGGGAACAA GGACAGGGG ATTTAAGCAG TCAAAAGGAA AAACATGTTA 360
 55 AGACCTAGA CTTGTATATT GACACACTTG TACCTTGTA GGCAGAGGAA TGTAATTAAA 420
 AAGCACTTAT TTGGCWNAAA AAAAAAAAAA AAAAAAAAAA C 461

60

(2) INFORMATION FOR SEQ ID NO: 154:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2388 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
10 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

5 GCCCACGCGT CCGAAAGCGG AGAACGCTGG TGGCCCTGTT GTGGAGTACG CTTTGGACTG 60
15 AGAAGCATCG AGGCTATAGG ACGCAGCTGT TGCCATGACG GCCCAGGGGG GCTGGTGGCT 120
AACCAGGGCC GCGCTTCAA GTGGGCCATT GAGCTAAGCG GGCCTGGAGG AGGCAGCAGG 180
20 GGTGGAAGTG ACCGGGGCAG TGGCCAGGGA GACTCGCTCT ACCCAGTCGG TTAATTGGAC 240
AAGCAAGTGC CTGATACCAG CGTGAAGAG ACAGACCGGA TCCTGGTGGA GAAGCGCTGC 300
TGGGACATCG CCTTGGGTCC CCTCAAACAG ATTCCCATGA ATCTCTTCAT CATGTACATG 360
25 GCAGGCAATA CTATCTOCAT CTTCCCTACT ATGATGGTGT GTATGATGGC CTGGCGACCC 420
ATTGAGGCAC TTATGGCCAT TTCAGCCACT TTCAAGATGT TAGAAAGTTC AAGCCAGAAG 480
30 TTTCTTCAGG GTTGGTCTA TCTCATTGGG AACCTGATGG GTTTGGCATT GGCTGTTTAC 540
AAGTGCCAGT CCATGGGACT GTTACCTACA CATGCATCGG ATTGGTTAGC CTTCATGAG 600
CCCCCTGAGA GAATGGAGTT CAGTGGTGGA GGACTGCTTT TGTGAACATG AGAAAGCAGC 660
35 GCCTGGTCCC TATGTATTTG GGTCTTATTT ACATCCTTCT TTAAGCCAG TGGCTCCTCA 720
GCATACTCTT AAATAATCA CTTATGTTAA AAAGAACCAA AAGACTCTTT TCTCCATGGT 780
40 GGGGTGACAG GTCTAGAAG GACAATGTC ATATTACGAC AAACACAAAG AAATAATACC 840
ATAACCCAG GCTGAAAATA ATGTAGAAAA CTTTATTTTT GTTCCAGTA CAGAGCAAAA 900
CAACAACAAA AAACATAAC TATGTAAACA AGAGAATAAC TGCTGCTAAA TCAAGAACTG 960
45 TTGCAGCATC TCCTTTCAAT AAATTAAATG GTTGAGAACA ATGCATAAAA AAAGTTGCAC 1020
AAGTTCCTTA TTTTCCTTAA TATTTCACTT CTATTTAATA CAAGCTGGGA CATAAAAATT 1080
50 CTGTTGGGGA TACCTGGGGG AAGATGTGAG AAATAATGC TGAATTCAGC TTATACATGA 1140
TGAAAAGAAA AACCAGACAA AAGGAGCACA TAAATATGCA TACAGTGTA CTGTTATTAT 1200
TTTAATACCC ACGATAAGG ATTTTGTGTA GCATGTTTAG GGGGAACGAG GATTGGTGGG 1260
55 ATCCTTGGGG CCACAGGAAT CTGAGGCAAC GGAAGATATA TAGAGTGATC GTCCCCCTGC 1320
CGAAGGAACC TGGCAYCTGT CAAGCAGATG CTGCAGTTCA AACTTCAGCT TTTAAGATAG 1380
60 ATAGCTATTG AAGGCAGAGG GTCAGCAGGA GGATGTGTAT TTCTAATCTA CCCTGGTAAA 1440

BEST AVAILABLE COPY

406

GTCATAGGTA AGACTCAAAA GCGGGATCTT ATTCAAAAGG CAGGTATTTT CTTTGTTC 1500
TGTCTTGAAA TAGCCCTTC CCTAAGGTG CATCTCTCA AGTTTTCAGT ATTGCTTTAT 1560
5 TTGCAGTGAT TAAAGAGAT GAGAGACTTT GGAGACAGAC AACGTAAGCA ACACATACAC 1620
ACATGAAATA CTCTAGACAG AGATGAATAT AAATCTGGCC TAATAACCAG TTTTCCATGT 1680
AACAGTGATT TTGTGTTTCG GGCTGAAGCA GTGGTTATAT TAAAAGCCAC TAATTCCCTT 1740
10 ATCCCTTTAA AAGATTTTCA CAATTCTCCA ACCACAAACA GCACTTCTAA AACTAACTTT 1800
ACTTCTGCC CATAATTTGT TCTACATGGA AAAAAAAT ATTACTTTGG CCAGGGGTGT 1860
15 GTGTAAATGT GGCAGAATTC CTAGGCAGGC TGACCTTTAC AGTATGGGCC TTTAAGATAC 1920
TGGATCCTGG TTGGGCAACA AGTGTACGC CTGAAGTTTC TGAAAACAAA TTAGAAGACT 1980
GTTGGCTTGG CTAATCTCGT AGTTCAGGC CAAGTTTCTG TAGTCAGAAT GAAGAATAAA 2040
20 ATTGAAAGAA AAAGGGGAA ATGCTTATAC TTGGCATTAA GTTGAATGCC TCAAGTCTTA 2100
ACTATGGCTT TGTAGATGAG GCAAAGATT TCTTAGTGGT AAAATTTCTT CAACAGGTCA 2160
25 ATGCCAATCT GTATGCCATT TTAGTAAAGT AGGTAAGGAG AGTAGCCGCT CAGTAACCTT 2220
GGCACTAAAG AAAGAGTGTG GCTCTAGAAC TTCCAATCCC ATTGCTAGAT GTGCCCTTTA 2280
AAAGATGGTC CAGTGCTTTC AGGGAAGGAT GTTAGCCAG TTTTCTAGT ATTTGTTCCT 2340
30 TAAGATTTTT TGACCTGTGC TTAATAAGAC GGACCGTGG GTCGACCC 2388

35

(2) INFORMATION FOR SEQ ID NO: 155:

40

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 642 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

50

55

60

AAAACAGACC ATTTAAAAC TCAGACAAGA TTATATTTAA TATATTAATT ACTAAAAGG 60
CACAAGATTA CACTGAACAT ATTAGCTACT AAAAAGGCAC TGCTAAGACA TTCAAGCAAA 120
TAGCTATTAC ACACTACTGC AGATTTTACA GTTTCTAAT TCTAACATAT GTTTGAAAAA 180
TCCGTAGTA TTCCAAAATA TATTTAATAA TGGAATATCT GCATTAATAT ACCATCCATG 240
TGTTTTTACC ATTGCTTA ATATTGAATA TACTGTTTAC CTCACACTAA AAAGAAAACC 300
AGAACCCTTA TTGTGATT TGGGAGTGA AGCTTCCATT TTTGTGTCAA AAATGAATCC 360
TGATCTTAT GGAAATCTCT GTTATTAAGA TATTCAAGA TGAGACAACA CTGAAGATCA 420
AATTGTGTTT AGTATCACTA TCTCTCTCC TCGTTTCTCT CTACTCCTC ATCCTCCAG 480

407

5 AATCTACCAG TTTATGGTAG AAAGATGGGA ACCTTATTG AATGTGTTTT TTTTTTCCA 540
TGATGTCCAA TTTTGTGTG GGAAAGGATT TCGATAAAAT TTTGTITTA ATTTTGGTAG 600
ATTTTTATCT ATACAAATTT AAATAAAATT ATGTTTTGTA AG 642

10

(2) INFORMATION FOR SEQ ID NO: 156:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1251 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

20 GCCGCTGCCC CTCCACGGAG TTGCTGATCA TCTGGGCTGT GATCCACAAA CCCGTTCTTT 60
TGTCCCTCCT AATATCAAAC AGTGGATTGC CTGCTGCAG AGGGGAAACT GCACGTTTAA 120
25 AGAGAAAATA TCACGGGCCG CTTCCACAA TGCAGTTGCT GTAGTCATCT ACAATAATAA 180
ATCCAAAGAG GAGCCAGTTA CCATGACTCA TCCAGGCACT GAGCATATTA TTGCTGTCAT 240
GATAACAGAA TTGAGGGGTA AGGATATTTT GAGTTATCTG GAGAAAACA TCTCTGTACA 300
30 AATGACAATA GCTGTTGGAA CTCGAATGCC ACCGAAGAAC TTCAGCCGTG GCTCTCTAGT 360
CTTCGTGTCA ATATCCTTTA TTGTTTGAT GATTATTTCT TCAGCATGGC TCATATTTCTA 420
35 CTTCAITCAG AAGATCAGGT ACACAAATGC ACGCGACAGG AACCAGCGTC GTCTCGGAGA 480
TGCAGCCAAG AAAGCCATCA GTAAATTGAC AACCAGGACA GTAAAGAAGG GTGACAAGGA 540
AACTGACCCA GACTTTGATC ATTGTGCAGT CTGCATAGAG AGCTATAAGC AGAATGATGT 600
40 CGTCCGAATT CTCCTTGCA AGCATGTTTT CCACAAATCC TCGTGGATC CCTGGCTTAG 660
TGAACATTGT ACCTGTCCTA TGTGCAAACT TAATATATTG AAGGCCCTGG GAATTGTGCC 720
45 GAATTTGCCA TGTA CTGATA ACGTAGCATT CGATATGGAA AGGCTCACCA GAACCCAAGC 780
TGTTAACCGA AGATCAGCCC TCGGCGACCT CGCGGGGAC AACTCCCTTG GCCTTGAGCC 840
ACTTCGAACT TCGGGGATCT CACCTCTTCC TCAGGATGGG GAGCTCACTC CGAGAACAGG 900
50 AGAAATCAAC ATTGCAGTAA CAAAAGAATG GTTTATTATT GCCAGTTTTG GCCTCCTCAG 960
TGCCCTCACA CTCTGCTACA TGATCATCAG AGCCACAGCT AGCTTGAATG CTAATGAGGT 1020
55 AGAATGGTTT TGAAGAAGAA AAAACCTGCT TTCTGACTGA TTTTGCCTTG AAGGAAAAAA 1080
GAACCTATTT TTGTGCATCA TTTACCAATC ATGCCACACA AGCATTTATT TTTAGTACAT 1140
TTTATTTTTT CATAAAATTG CTAATGCCAA AGCTTTGTAT TAAAAGAAAT AAATAATAAA 1200
60

ATAAAAAAAAA AAAAACCCCG GGGGGGGCCC GGTOCCCAAT TGGCCCTATG G

1251

5

(2) INFORMATION FOR SEQ ID NO: 157:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 2127 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

20

COGGCGGGAG AGGGAAGCTG CAGCGAGAGG CGCGGATCTC AGCGCGGGAG CAGTGCTTCT	60
GCGGCAGGCC CCTGAGGGAG GGAGCTGTCA GCCAGGAAA ACCGAGAACA CCATCACCAT	120
GACAACCACT CACCAGCCTC AGGACAGATA CAAAGCTGTC TGGCTTATCT TCTTCATGCT	180
GGGTCTGGGA ACGCTGCTCC CGTGGAAATT TTTTCATGAC GCCACTCAGT ATTTACAAA	240
COGCCTGGAC ATGTCCAGA ATGTGTCTT GGTCACTGCT GAACTGAGCA AGGACGCCCA	300
GGCGTCAGCG CCCCCTGCAG CACCCTTGCC TGAGCGGAAC TCTCTCAGTG CCATCTTCAA	360
CAATGTCATG ACCCTATGTG CCATGCTGCC CCGCTGTGA TTCACTTACC TCAACTCCTT	420
CCTGCATCAG AGGATCCCC AGTCCGTACG GATCCTGGGC AGCCTGGTGG CCATCCTGCT	480
GGTGTCTCTG ATCACTGCCA TCCTGGTGAA GGTGCAGCTG GATGCTCTGC CCTTCTTTGT	540
CATCACCATG ATCAAGATCG TGCTCATTAA TTCATTGGT GCCATCCTGC AGGGCAGCCT	600
GTTTGGTCTG GCTGGCCTTC TGCCTGCCAG CTRACCGGC CCCCATCATG AGTGGCCAGG	660
GCCTAGCAGG CTCTTTTGCC TCCGTGGCCA TGATCTGCGC TATTGCCAGT GGCTCGGAGC	720
TATCAGAAAG TGCTTGGC TACTTTATCA CAGCCTGTGC TGTKATCAIT TTGACCATCA	780
TCTGTACCT GGGCTGCCC CGCCTGGAAT TCTACCGCTA CTACCAGCAG CTCAAGCTTG	840
AAGGACCCGG GGAGCAGGAG ACCAAGTTGG ACCTCATTAG CAAAGGAGAG GAGCCAAGAG	900
CAGGCAAAGA GGAATCTGGA GTTTCAGTCT CCAACTCTCA GCCACCAAT GAAAGCCACT	960
CTATCAAAGC CATCTGAAA AATATCTCAG TCCTGGCTTT CTCTGTCTGC TTCATCTTCA	1020
CTATCACCAT TGGGATGTTT CCAGCCGTGA CTGTTGAGGT CAAGTCCAGC ATGCGAGGCA	1080
GCAGCACCTG GGAACGTTAC TTCATTCTCT GTCTCTGTTT CTGACTTTC AATATCTTTG	1140
ACTGGTTGGG CCGGAGCCTC ACAGCTGTAT TCATGTGGCC TGGGAAGGAC AGCGCTGGC	1200
TGCCAAGCTG GNTGCTGGCC CGGCTGGTGT TTGTGCCACT GCTGCTGCTG TGCAACATTA	1260
AGCCCCCGCG CTACCTGACT GTGGTCTTCG AGCAGGATGC CTGGTTCATC TTCTTCATGG	1320
CTGCCTTTGC CTCTCCAAC GGCTACCTCG CCAGCCTCTG CATGTGCTTC GGGCCCAAGA	1380

AAGTGAAGCC AGCTGAGGCA GAGACCGCAG AGCCATCATG GCCTTCTTCC TGTGTCTGGG 1440
 TCTGGCACTG GGGGCTGTTT TCTCCTTCCT GTTCCGGGCA ATTGTGTGAC AAAGGATGGA 1500
 5 CAGAAGGACT GCCTGCCTCC CTCCTGTCT GCCTCCTGCC CCTTCCTTCT GCCAGGGGTG 1560
 ATCCTGAGTG GTCTGGCGGT TTTTCTTCT AACTGACTTC TGCTTTCCAC GCGGTGTGCT 1620
 10 GGGCCCGGAT CTCCAGGCC TGGGAGGGA GCCTCTGGAC GGACAGTGGG GACATTGTGG 1680
 GTTTGGGGCT CAGAGTCGAG GGACGGGGTG TAGCCTCGGC ATTGCTTGA GTTCTCCAC 1740
 TCTTGCTCT GACTGATCCC TGCTGTGCA GGCCAGTGA GGCTCTTGG CTGGAGAAC 1800
 15 ACGTGTGCT CTGTGTATGT GTCTGTGTGT CTGCGTCCGT GTCTGTGAGA CTGTCTGCCT 1860
 GTCTGGGGT GGCTAGGAGC TGGGTCTGAC CGTGTATGG TTTGACCTGA TATACTCCAT 1920
 20 TCTCCCTGC GCCTCTCCT CTGTGTCTC TCCATGTCCC CCTCCCACT CCCCATGCCC 1980
 AGTTCTTACC CATCATGCAC CCTGTACAGT TGCCACGTA CTGCCTTTT TAAAAATATA 2040
 TTTGACAGAA ACCAGGTGCC TTCAGAGGCT CTCGTATTA AATAAACCTT TCTGTTTTT 2100
 25 TTCTCCATGG AAAAAAAAAA AAAAAA 2127

30

(2) INFORMATION FOR SEQ ID NO: 158:

(i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 1625 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

40

CAAAAGATCT ATAATCAGGA CATGTATTAT GTAAGTTGGA CAANAAAAAT TCTTCCCTT 60
 TATGTCCACC CTTCTATGA TTGCAAGACA AAATTCCCT CTTTACCTC ATCCCTATAA 120
 45 CATGGGAGGC TGAGAAAAAT GAGGGAGAT GGAACCAGAT ACAAGGAGAT CCAATAAGAG 180
 AAGCTTATTT AAATATTGTG AATAAAGGA AGAMCCAAAG CATTTTTTTA AGTGGGGAAT 240
 CCTTTGAAC AGTTATTATT TATCCATATT ATTAAYAACA TCTTTCTGA CAAAATCCAT 300
 50 CAGATGAAGT GTAAATGGAT AATCTTTTAA TGGATCTAAA CTAGAAAGT TTCCTTACT 360
 GTTCATGTCC GTGTCCAGA ATTGTGAAAT GGTGTGTGGT TTTGCTTTCC AAGTCTTCT 420
 55 CTGCCTCTC TTAATCTCT AATCCATGT CTACAGAAG AATGAGAAAT TTCTTTCTTA 480
 CTTGAGTATC ATGCTCTAAA AAAGTGGCT TCAGTCACAG AAACGCTGGC TCTCCTGTGC 540
 TTATATTGAA GCCAACTGCC TTTAATCTT GGGCCCTCTT ATATTTTAA GTGCAAAAT 600
 60

	TTGAAGTCTC AGTCACCAGA CACAGGTTCT ATACAATTAA TGATGAGCTG GAGAAGTAAT	660
	ATGTAGCTAA TTTTTCAAAA GCATTGAATA TACTTTCCGG AAAGAAAACA GAAATTAAAT	720
5	ATTGCCACAT CTGCCAGAA TCCCATCTGA CACCTTAACT TTGTCAGGTT TCCTACAACT	780
	TGCTAATCAA GTTTTATACA TTCTAAATCT CCCAGTTTC TTGGGGCTG GAAGATGCAA	840
10	CTTCCATTTA ATAGAACTT TGAAATCTTG GGGTAAGGGA GCAGTGGGG GACTAGGGAG	900
	AAGGATAAGA AATAGAATTA TTGAAAAGCC CCCACCAGG ACCTTCCTGG CCAGAATATG	960
	CAGAGTAATT CCGTCTGGCT TCACCTTTGA AAGTCCCTCG AACTATGCA GATGAAACTG	1020
15	AGTCTGTTTT TGATATTGTC AGATGTATTC TACCTTGAA GTCCCNACAC CTAAACTGGA	1080
	ATTCTGTAT TTACATCTCC TCCACTGTCC CCCACACCAC CCTCAATTC CTGCTGCCCC	1140
20	TGCTAATGTT AAGCATTTTT CTCCTGTTAT CATCAGGTT ACATTAAAAM CAGRTACTTA	1200
	CAAACTGACT TGAAGCACAG ATACTTTTAC GAATGTGATA AAATATTTTC TTAAGAAAAG	1260
	GAAAGAGGAT GTGGGTCAA TAAACACCG CATGGATGTT GATTGGTGAA TACTGGTGTA	1320
25	AGAAAAGGGA GCTCAGGAAT TTTTATTACT GTATTTGTAA ATGAGTTTGA AGGAATTTGT	1380
	AAATGCCACT GGTACATTTT TAAGGTGACA CATTGTCTCC TTATAAGTT ATTAAAAATT	1440
30	ACAGGGTAAG CTTAAATGAC GTTGCCAGT AGTTTACTT TATATAATCA ATATTGATAT	1500
	TGTTGCTGAA CTATGTAAT TTATGATGCA TTTTTCAGTC CCTTTTCAGA GCAAATGCTT	1560
	TTGCAATGGT AGTAATGTTT AGTTTAAATT GACTTAATAA ATTTTACCT GAGCAAAAAA	1620
35	AAAAA	1625

40 (2) INFORMATION FOR SEQ ID NO: 159:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1687 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

50	CGGGGTACC AGTTATTAGA GGAAGTAACA CAAGGGGATA TGAGTGCAGC AGACACATTT	60
	CTGTCCGATC TGCCAAGGGA TGATATCTAT GTGTCAGATG TTGAGGACGA CGGTGATGAC	120
	ACATCTCTGG ATAGTGACCT GGATCCAGAG GAGCTGGCAG GAGTCAGGGG ACATCAGGGT	180
55	CTAAGGGACC AAAAGCGTAT GCGACTTACT GAAGTGCAAG ATGATAAAGA GGAGGAGGAG	240
	GAGGAGAATC CACTGCTGGT ACCACTGGAG GAAAAGGCAG TACTGCAGGA AGAACAAGCC	300
60	AACCTGTGGT TCTCAAAGG CAGCTTTGCT GGGNATCGAG GACGATGCCG ATGAAGGCC	360

TGGAGATCAG TCAGGCCAG CTGTTATTTG AGAACCGYG GAAGGACGG CAGCAGCAGC 420
 AGAAGCAGCA GCTGCCACAG ACACCCCTT CCTGTTTGAA GACTGAGATA ATGTCTCCCC 480
 5 TGTACCAAGA TGAAGCCCT AAGGNAACAG AGGCTTCTTC GGGGACAGAA GCTGCCACTG 540
 GCCTTGAAGG GGAAGAAAAG GATGGCATCT CAGACAGTGA TAGCAGTACT AGCARTGAGG 600
 10 AAGAAGAGAG CTGGGAACCC TCCGTGGTAA GAAGCGAASC GTGGGCCTAA AGTCAGATGA 660
 TGACGGGTTT GAGATAGTGC CTATTGAGGA CCCAGCGAAA CATCGGATAC TGGACCCCGA 720
 AGGCCTTGCT CTAGGTGCTG TTATTGCCTC TTCCAAAAG GCCAAGAGAG ACCTCATAGA 780
 15 TAACTCCTTC AACCGGTACA CATTTAATGA GGATGAGGG GAGCTTCCGG AGTGGTTTGT 840
 GCAAGAGGAA AAGCAGCACC GGATACGACA GTTGCTGTGT GGTAAAGAAG AGGTGGAGCA 900
 20 TTACCGGAAA CGCTGGCGGG AAATCAATGC ACGTCCCATC AAGAAGGTGG CTGAGGCTAA 960
 GGCTAGAAAG AAAAGGAGGA TGCTGAAGAG GCTGGAGCAG ACCAGGAAGA AGGCAGAAGC 1020
 CGTGGTGAAC ACAGTGGACA TCTNCAGAAC GAGAGAAAGT GGCACAGCTG CGAAGTCTCT 1080
 25 ACAAGAAGGC TGGGCTTGGC AAGGAGAAAC GCCATGTAC CTACGTTGTA GCCAAAAAG 1140
 GTGTGGGCG CAAAGTGGC CGGCCAGCTG GAGTCAGAG TCATTTCAAG GTGGTGGACT 1200
 30 CAAGGATGAA GAAGGACCAA AGAGCACAGC AACGTAAGGA ACAAAGAAA AAACACAAAC 1260
 GGAAGTAAGC AGAGCTGCCA GGCTCCAGG AGAGCATGGG GACTAGGAGG AAGGTGTGG 1320
 CATGGCTCAG TCTGGCCCC TTGATTACCG GCCTAGCCCC TGCTCACATC ACAGCTGTCT 1380
 35 GAAGAACAGT GAGGTGGAGT GCCTAGAACT CCCGTGGTGG TCCTGAGCAG AGAGGAGGAT 1440
 GTCTCTCTGC CTGCCTGAAG GTCTCCCATG AAAACACTGC TGAAGTGTGT TGACACTCAT 1500
 40 GACCTTTTTT TTAACCGTT AAAGGGAAGT TCGGTGTGG AGCGATACTC AATGTAGTCA 1560
 GTCTACACCT GGACGTGTGG GCACTTAAG CCTCCCCAC CCCCATCCTA TTCCTRAATA 1620
 45 AAACCAGGAT AATGGAARAA AAAAAAAAAA AAAAAAAG GGGGGGCCN TAAAGGNCC 1680
 CANNITT 1687

50

(2) INFORMATION FOR SEQ ID NO: 160:

(i) SEQUENCE CHARACTERISTICS:

55

- (A) LENGTH: 1842 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

	GGATGACAGA TTGCGACANA GATTTGTGAC CCTTCCTGCT GAACTTCAGA GGGAGCTGAA	60
	ANCAGCGTAT GATCAAAGAC AAAGGCAGGG CGAGAACAGC ACTCACCAGC AGTCAGCCAG	120
5	CGCATCTGTG CCCCAGAGAAT CCTTTACTTC ATCTAAAGGC AGCAGTGAAA GAAAAGAAAA	180
	GAAACAAGAA GAAAAAACC ATTGGTTCAC CAAAAGGAT TCAGAGTCCT TTGAATAACA	240
10	AGCTGCTTAA CAGTCCTGCA AAAACTCTGC CAGGGGCTG TGGCAGTCCC CAGAAGTAA	300
	TTGATGGGTT TCTAAACAT GAAGGACCTC CTGCAGAGAA ACCCCTGGAA GAACTCTCTG	360
	CTTCTACTTC AGGTGTGCCA GGCCTTTCTA GTTTGCAGTC TGACCCAGCT GGCTGTGTGA	420
15	GACCTCCAGC ACCCAATCTA GCTGGAGCTG TTGAATTCAA TGATGTGAAG ACCTTGCTCA	480
	GAGAATGGAT AACTACAATT TCAGATCCAA TGAAGAAGA CATTCCTCAA GTTGTGAAAT	540
20	ACTGTACTGA TCTAATAGAA GAAAAGATT TGGAAAACT GGATCTAGTT ATAAAAACA	600
	TGAAAAGGCT GATGCAGCAA TCGGTGGAAT CGGTTTGGAA TATGGCATT TACTTTATTC	660
	TTGACAATGT CCAGGTGGTT TTACAACAAA CTTATGGAAG CACATTAAAA GTTACATAAA	720
25	TATTACCAGA GAGCCTGATG CTCTCTGATA GCTGTGCCAT AAGTGCTTGT GAGGTATTTG	780
	CAAAGTGCAT GATAGTAATG CTCGGAGTTT TTATAATTTT AAATTTCTTT TAAAGCAAGT	840
30	GTTTGTACA TTTCTTTTCA AAAAGTGCCA AATTTGTCAG TATTGCATGT AAATAATTGT	900
	GTTAATTATT TTACTGTAGC ATAGATTCTA TTACAAAAT GTTTGTTTAT AAAGTTTAT	960
	GGATTTTAC AGTGAAGTGT TTACAGTTGT TTAATAAAGA ACTGTATGTA TATTTGGTAC	1020
35	RGGCTCCTTT TKGIGAAYCC TTA AAAACTC AACTCTAGGA RGCAACTACT GTTTATTATA	1080
	CTAAARGGCT GAAAAMCCTC CAGGCAGAC TGCTAAGCTC TGAAATYCCT GAGAGGCTC	1140
40	AGACCGGGAT TCTACTTGT CCAAGAAAGG GTAAAGCTTC TAAACCATCT TATTCTGTG	1200
	TCCAAGCATG AACACAGGAG CATGTYAAGA AAATCTTTAC TACTTTCTYC CATGCCGAGA	1260
	AATCTACATA TTTTGAATTA GAAACACCT CACACCCACT TGAAGATTTT TTTCTGGGA	1320
45	ACATTATGTC CCGTAGATCA GAGGTGGTGT TGTCTTTTG CTTCTACTGG CCATTGAGAA	1380
	ACTTTGATGA TAAAAAGAA CGGTATAGAT TTTTCAAACG TATATAAAT ATTTTATGT	1440
50	TATATGTTAT GCCATAACTT TAAAAAATAA ATAGTTTAAA ATTCTATGCT AGTGGATATT	1500
	TGGAACTTT TCTCAAACA AACACCCAC ACTGACTTCA GCAAAACCT AAAACTAGCT	1560
	ACAGATTACT ACTACGAATG AATCATYAAG TTTTGTGTCT GCAACAATTT AGAAGCACTA	1620
55	AGCCCAAATA TCAGGAAATG TGTGTATGAT GGAATTTTCT AGGACAAAAC AGATCAAGAT	1680
	TAAACAGGA TCAAGGATTA ATGGTATAAA AATGGTCTAC TAAACAGGA TCAAGGATTA	1740
60	AAACAGGATC AAGGATTAAT GGTATAAAAA TCTCTACTGG TTACCGGGTG GCNGGGCCAT	1800

ACAGGGTAGT GGTGGATGGA TAGTTTAGTT TGGNAAGGGT AA

1842

5

(2) INFORMATION FOR SEQ ID NO: 161:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 770 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

15 GGCAAGAGCC CTATGCTGTT CTTGTGATAA TGAGTGAGTC TCACAAGATC TGGTGGTGTT 60
ATAGGCATCT GGCATTTCCT CTGCTGACGC TCATTCTCTA TCCTGCCACC CTGGGAAGAA 120
20 GTGTCTTCG TCATGATTGT AAGTTTCTG AGGCCTCCCC AGCTATGTAG AACTGTGAGC 180
CAATTAAACC TCTTTTCTCT ATAAATTATC CAGTCTTATA TATTTCTTCA TAGCAGTGTC 240
AGAACAGATA ATACCGTAAA TTGTATCAC AGAGAGTGGG GTGTGCTAT AAACACATCT 300
25 GAAATGTTA AAGCAAATTT GGAAGTGGT AACAGGCAA GGCTGGAACA GTTGAAGAA 360
CAGTTAAGAA GAAGACAGGA AAATATGAGA AATCTTGAAA CTCTAGAG TCTTAAAGT 420
30 CTCAGAAGAC ATGAAGATGT GGAAGCTTT GGAAGTCTCT AGAGACTTGT TTGAATGGCT 480
TTGACCAAAA TGCTGATAGT GATATGGACA ATGAAGTCCA GGCTGAGCTT ATCCAGACAG 540
ACATAAGAAG CTCGCTGGA ACTTGAGTAA AGATCACTCT TGCTAGGCAA AGAGACTGGT 600
35 GGCTTTTCTT CCTCTGCCCT AGAGATCTGT GGAAATCTGA ACCTGAGAGA GATGATTTAG 660
GGTATCTGGC AGAAGAAATA TCTAAGCGGC AAAACCTTCM AGAGGAAGCA GAGCATAAAC 720
40 GTTTGAAAAA TTGCAGCCT GACNATGGGA GACCAAAGTT AAACCAATT 770

45 (2) INFORMATION FOR SEQ ID NO: 162:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 519 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

55 GAATTCGGCA CGAGCTGAGA GGCACAGGAG CAACAGCCAG TGCCCCCTGC AGAGGACCAC 60
TGGGGTCACA GACTTCARAC CTGATGACCT GGGCTCAGAT CCCAGCTCTG CACCTACCAG 120
60 CCGTGTGACA AGGTGTCCTC TCTGAGCCTC AGTCACACAC TGCCTTAACG GTTGGGCCTC 180

414

ATGGAGCTGT TTGTGAAGGT TAAATGGGAA GACATAAAGC ACTTAGCCCA GAGCCAAGGA 240
CATGCTGAAT AGGATAATGG TGGCCTCCTT TGGCGCTGTG CTGGTGCAAG TGTGCCGAGG 300
5 AAYTGGGCAG GGGTGACAGA TACCTCTTCT AACCTAGTTC CTTTCCAAGA ACCTAATTGG 360
TGTCTCTCCC TCCCCCAGGC AATTGGAAGG AGGAGGCTGG GCCCCAGCCC CAGAATACGG 420
GAGGTTTCTC ACCGTGGTAG GGAAATGCT GGGTTGGGGG TGTGGGCAAC CACAGTGATC 480
10 GTCTCTCTGC AGGACGGATG AGGCTTTGCT GACAGAGGC 519

15

(2) INFORMATION FOR SEQ ID NO: 163:

(i) SEQUENCE CHARACTERISTICS:

20

- (A) LENGTH: 753 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

30

35

40

45

50

GGCAGGAGCG GCACGAGCAG CCAGTTGCTG ACTGGCACAT GGCTCCAGC GTCCCGGCTG 60
GTGGGCACAC TAGAGCCGA GGGATCTTCT TAATTGGTAA ATGGATCTT GAAGCTTCAC 120
TGTTTAAATC TTTTCAGTGG CTCCCTTTG TACTTAGAAA AAAATGCAAC TTCTTCTGCT 180
GGGACTCATC CGCTCACAGC CTCCCCCTCC ACCCTCTCTC TGCTCATGC TCTGCCCCTG 240
CCTGCCATGC CTCGATACT CACCTTTTGT ACCCCAGCAC CCGTGCCCTC TGCCCCCGA 300
TCTTTGCCCTG GCTGGTTGCT CCTCACTCAG TGTTCAGGAC AAATGCTCCT GGCCCTACCC 360
CATCTAGCCA GTCTAGCCCG GTCTTCCCTG TCTTCCCTGT TTCATTATG GCTCTTATG 420
TTTGTTWACT TGTGTCTGT TGACTTTTAA CTCTCTCAGT CCCCCTGGA ATGCAAGCGA 480
TCTCCCAAGC TCCTAGAATT GTTCTGCCT CTTCACAGGC CCTTACGCTG TGTGTGCTCG 540
TGCCGAATTC GGCACGAGG TATGTGCACT TGCTGGTATG TATGTAGGTG TTTGCTAACA 600
CATACGTGCA CACGCAGAAT GCTTCCAGGG GACTGCACAG CCTCTAGTTC GCAGCCCCCA 660
CCCCCTCCCTT TGSCCCTGCA CTCTCCCTC TCTGAGCTGC ATTGCGATGA AAGGGTGCA 720
50 GGTTCCTGAN CCCGCNAGCG NCACCTCCTG GGA 753

55

(2) INFORMATION FOR SEQ ID NO: 164:

(i) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 1400 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double

415

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

5	GGCACAGTTT ATTAATACCT ATTATGGGA AGTCACTTTG GTTGGCATTG AAAATTACAT	60
	CATCTTTAAA GCAGTATTTG TCCCCAGATG GACTCATCAC TAGCAAAGAC TAGGTTTCATT	120
	GGAAGGCATA GGGTGAGAGA ATGGGAAGAT GRAGTGGAGG CGGGTTGTTA AAGTGCTGTC	180
10	AGTGAGTGAT TTTGTCTACT TGAACAATGG TCCATGTTTG GGGGCATATT GTGTTTCATA	240
	AGAAGTGAAA GGTATTTGCA AAGTAAGCTA CAAATGACCC ATAAATCTGT TAACAACAGT	300
15	CCTTAATATG CAAAGATGAA AAACAAGCAT TACTGCTACC CAAAGGGAAC TGGTGCTTGG	360
	TGATGTGCAG ATGGGGCTGT TGGTTAAGAG AGCTATTACA GGTTTTCTCT CTTAGGTTTC	420
	ATAGGAGGTA GTTACTGAGA TGAGATTGTT TTATCTTTT GAATACAGAT CTCTGTCTTT	480
20	GAGTTAGTTC TGAGGATGGG AGTAATAAAG GAGTTTTTGG TTTTTTGTG TGTGTTTGG	540
	TTTTGGCTCC TTAGTAATAC TCCTCTGACA TTTATTTCTA TTATTCTTCA AAGAAAGGAA	600
25	ACCAACTGAA ATGTTTGCTT TAACAACAT TTTAATAAGT TCTCTGGGTT TTTTTTCCC	660
	CTTTTAAAAA AATTAGCATA TACCATAGCA ATAAAGAAC TAATGTTAAC TATTGTATGC	720
	TACAACTTAA GTGATTTTTC TAAAGAAGCA CAATGTCATT GRAAGTATTA TTGAAAAGGA	780
30	TCATAGTCAC ATTGAATTTG TGAAGGCCAA AGAAATTGAA GGGAGTGATA TTTTCATTTT	840
	ATGATATCA CATATTTAGT AAATTTTGTG TACAAGAATA CCAGGCAGAG TGTTTTACCC	900
35	ATGGAAACAG GTTTCAGATT ACTTTGTTTT TACTGTTAGA GTCTCAAGTT TAGAAATGCT	960
	AACACTTAAA TCAGTTTTTT TCTCACTATA CTGAAGATT GTTAATATTT TGATATCTTC	1020
	CTAGCTTGAT GGAATTTAAA CATATCTTCA GATCTGTGAC AGTGACAGCC AATAGGACTG	1080
40	ATAATATTAG CTTCAAACCA ATAATATCCA GGGTTAAAAT AAAAATCATA GTGAAAGTAC	1140
	GATTGTAAAA TTATGCTATA TTAACTTTGA AGTCTGTAAT AACTTGACAT CAAAATGTTA	1200
45	TGTAATTACC ATAAATAATG GCTAGCGAGA ACATCTTTGG AAATCTCAA ATTACCTTTC	1260
	TTACTACACT GTTTCAGAAA TGAATGTAGA AATGATCCTG TTAGCTTTCT GAATGTTCTG	1320
	TGGTTGAATG TGTTTTGTCT TAAATAAAGC TTTTGGTATT TGTTTAAATW ACAAAAAAAA	1380
50	AAAAAAAAAA AAAAAGTGA	1400

55

(2) INFORMATION FOR SEQ ID NO: 165:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2153 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

5	CAGGGCTCAG GGCCTCTGGT GGCTCTGGCC CAGACAGTAT TTGCAGTTCT TGTGCTATGG	60
	GTGGGAGTCT TCTTCCTCAA GTTTCGGCAG CTGTGCTGTG NCTGGATGGG CTGCTCCTCC	120
10	CAGGGCTCAA GGGCTGTGGT CCGCTCAGGG TCTCATTTCC CCAGGCCAAG TTCAAGGCAG	180
	CAGCCCTTTG TGAGGCGCTC TTGGCCCTGG GCTGGAGGGA GAACTTTAAG CTTTTTTGCT	240
15	CACAGGGACG TGGTATGGGC CCTGGGTGCA GGTGCCACACA TTCTGCTAAT GAGAGCTTTG	300
	TCTGATCAGT CCTGGGTCCA TCAGTTTGTG CATGTGTCCG GCTGCCAGCC CGTCCCTTGG	360
	GATCCTTCCC CTGGGGTGTA GCCTTGTTCA TTAGTATATA CTCATTCCCTT CATGCTTTCC	420
20	TCAGCAGAAC ACTTCCACTT CTGAGGTGAG CTTTTGCCCC RTGCCCTTCC TCCACAGGTG	480
	TTGCCPTTTT ATAAAGACCT GATAGCAGAA TAAATTGGTG TTTCCCTGTT GACCCAGCAC	540
	CATTTCTGTG GGCCTAGAAT ATGGCCCTCA ACCCTTAGAG TGGGGCAGTG AGGGCTTGAG	600
25	GAGTGACCTT TCCTTTCTCA TGGTTTGTAGT CATTTTGGCT GCCAGCCCTT AATGGCACAG	660
	ATCTGCTGCT TCTAACAGAT GGCCAGGAGG TGACACOGAT TTCAGCCATT GCCAAGGTTA	720
30	GCACCCCTCT CTTTGAGCCT AGGGCCACAC TGTTCATTGT CACTTTAGGC AAGTGCCCTG	780
	TTGGCTTTAA AGGTAAGCCT GCCAGCTGTG AGAAGCCTTG GTAACGTATG GACTCATTTT	840
35	CTGGTCTTA AAGATGCAGC CTCTTAAGGG CTCCTTGATG GATGCCATCT CTCCTAGCCC	900
	CCAGCCCTGG TGCCACTGGT GGGCAGGTTT CCATTCTTTG GGGCTGGGAG GGACAGCTTG	960
	CCTGTTCTG GTCACAAATT ACAGTCTTCT CTCCTGTACC ATTCTGTGGC TTCAGCATGG	1020
40	GGGCAGTAGC CTTTCATTAG TGTAGATAGT CATTCCTTGG TAGGGTGGAG GGTAAGACAT	1080
	AGGGTCTGGA ACTGTTTGGG ACCTTTTGGG GATGTCTGT GCCTCCCA GATCTTGATT	1140
45	CTGGGAGGAG AGGCTGCCGC ATTCTGTGTC TCCTCACAGC GAGCAAAGCT GCACCCACTT	1200
	ACATTAGTA TTTTCTGGC ACTACAAAGA GTGGGAAGGC CTGGGATTG CTGCTGCTCC	1260
	CTTAGAGCAG GGCCCTTCTT TTCAGCACTT TGGACACCTG GAGACCCAGC CCTGTTATTT	1320
50	AATGGTAGTG GCAAAGTGTG TGTGCATACT GTCTGCCACT GCTTCTTCCC TGCCCCATGC	1380
	CAGAGAGCCC TGTCCCTGCC AGGCCAGCC TTCTTAGCCC CAACTTGGGA ACAAGTGCA	1440
55	ACATGGGATC ATGGGTTGGG GTGCTCAGGT GAGCCCTCTC TATAGTGCTT CCCTGGGCA	1500
	AGCTGACACC AGCCCTGAG GGTGGGGTGG GACGGGTGGT GCTTAAAGA GGAAGGGGAC	1560
	CAGTGTAGCA ACTTGCCAGG GACCCACCC CTCCTCTCT GGGCTGTGC AGTGAGCATG	1620
60	GGGATTCCA TCAAGGGGCC TGGCACCTGT GCTAGTTACG TAGCCGCTGN TCACGCGCTC	1680

417

ACTCCTGACC ACATGCACGT TCCCTAGATG CAGACTGCTT TGAAC TTAA AGCTGTACAA 1740
 TTTGGTTATG TTTGTGCTGA CTTAAAATAT ATTTTAATGA GGAAAAATA ATGGAGAACC 1800
 5 CTGGAAGGA CTTGGTCTT TGTCTCTCG GGAAGCTGA AGCCCTCGG TTCTGGGAAT 1860
 CGCTCTCTGC TGCTCTTTCC TGAAGCTAA GCCTGTCTCC ACCGCCCGAG GCCTGCGCCG 1920
 10 GTGCTCCCGC CGCAGTTGCG TTTGCTTTGG ACCTTGCGTG CGGGGGAGGG GGTGCTCGGT 1980
 CCGAGCCCGC TCCTTTCTGT ACACCTAGCG CTGCCCGCCC CGCTTGCTG TGAGGTCTGT 2040
 TATGTCAAAA ATAAAGCCGC TAGAAACGGA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 2100
 15 AAACCTGAGG GGGGGCCCGT ACCCAATTAA CCNNTATGA TCTATAAAGC GTC 2153

20
 (2) INFORMATION FOR SEQ ID NO: 166:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1251 base pairs
 25 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:
 30 GCCCAGCGGT CCGCCACGC GTCCGCGGT GCGAGTATG GGGCGTGAT GGCCATGGAG 60
 GGCTACTGGC GCTTCCTGGC GCTGCTGGG TCGGCACTGC TCGTCGGCTT CCTGTGGTG 120
 35 ATCTTCGCCC TCGTCTGGGT CCTCCACTAC CGAGAGGGG TTGGCTGGGA TGGAGCGCA 180
 CTAGAGTTTA ACTGGCACC AGTGCTCATG GTCACCGCT TCGTCTCAT CCAGGGCATC 240
 40 GCCATCATCG TCTACAGACT GCCGTGGACC TGAATGCA GCAAGCTCCT GATGAAATCC 300
 ATCCATGCAG GGTAAATGC AGTTGCTGCC ATTCTTGCAA TTATCTCTGT GTGGCCGTG 360
 TTTGAGAACC ACAATGTTAA CAATATAGCC AATATGTACA GTCTGCACAG CTGGGTGGA 420
 45 CTGATAGCTG TCATATGCTA TTGTTACAG CTTCTTTCAG GTTTTTCAGT CTTTCTGCTT 480
 CCATGGGCTC CGCTTCTCT COGAGCATTT CTCATGCCCA TACATGTTTA TTCTGGAATT 540
 50 GTCATCTTTG GAACAGTGAT TGCAACAGCA CTTATGGGAT TGACAGAGAA ACTGATTTT 600
 TCCCTGAGAG ATCCTGCATA CAGTACATTC CCGCAGAAG GTGTTTTCGT AAATACGCTT 660
 GGCCTTCTGA TCCTGGTGT CGGGCCCTC ATTTTTTGA TAGTCACCAG ACCGCAATGG 720
 55 AAACGTCTTA AGGAGCCAAA TTCTACCATT CTTATCCAA ATGAGGCAC TGAACAGGGA 780
 GCAAGAGGTT CCATGCCAGC CTACTCTGGC AACACATGG ACAAATCAGA TTCAGAGTTA 840
 AACAGTGAAG TAGCAGCAAG GAAAAGAAC TTAGCTCTGG ATGAGGCTGG GCAGAGATCT 900
 60

418

ACCATGTAAA ATGTTGTAGA GATAGAGCCA TATAACGTCA CGTTTCAAAA CTAGCTCTAC 960
AGTTTTGCTT CTCTATTAG CCATATGATA ATTGGGCTAT GTAGTATCAA TATTTACTTT 1020
5 AATCACAAAG GATGGTTTCT TGAAATAATT TGTATTGATT GAGGCCTATG AACTGACCTG 1080
AATTGGAAG GATGTGATTA ATATAAATA TAGCAGATAT AAATTGTGGT TATGTTACCT 1140
TTATCTTGTT GAGGACCACA ACATTAGCAC GGTGCCTTGT GCAKAATAGA TACTCAATAT 1200
10 GTGAATATGT GTCTACTAGT AGTTAATTGG ATAACTGGC AGCATCCCTG A 1251

15

(2) INFORMATION FOR SEQ ID NO: 167:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 882 base pairs
20 (B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

25 GACSMCTAG AACTATGGTC CCCCCGGACT GCAGGAATC GGCACAGCG CTGCGGGCGC 60
GAGGTGAGGG GCGGAGGTT CCAGCAGGA TGCCCCGGCT CTGCAGGAAG CTGAAGTGAG 120
30 AGGCCCCGAG AGGGCCAGC CCGCCCGGG CAGGATGACC AAGGCCCGGC TGTTCCGGCT 180
GTGGCTGGTG CTGGGGTCGG TGTTCATGAT CCTGCTGATC ATCGTGACT GGGACAGCGC 240
AGGCGCCGCG CACTTCTACT TGCACAGTC CTTCTCTAGG CCGCACACGG GGCCGCCGCT 300
35 GCCCACGCC GGGCCGACA GGGACAGGA GCTCAGGCC GAYTCOGATG TCGACGAKTT 360
TCTGGACAAK TTTCTCAGT CTGGCGTGAA GCAGAGTGAC YTTCCAGAA AGGAGACGGA 420
40 GCAGCCGCCT GCGCCGGGA GCATGGAGGA GAGCGTGAGA RGCTACGACT GGTCCCCGCG 480
CGAMCCCCG CGCACCCAGA CCAGGGCGG CAGCARGCG ANCGAGGAR CGTGCTGCGG 540
GGCTTCTGCG CCAAYTOCAG CCTGGCCTTC CCCACCAAG AGCGCGCATT CRAAGACATC 600
45 CCCAACTCG AGCTGAGCCA CTGATCGTG GACGACGGC ACGGGGCCAT CTACTGCTAC 660
GTGCCCAAG TGCCCTGCAC CAACTGGAAG CGCGTRATGA TCGTGCTGAG CGGAAGCTGT 720
50 GCACCGCGTG CGCTACCGC GACCCGYTGC GNTCCCGGC GAGCACGTGC ACAACGCCAG 780
CGCGCACTGA CTCAACAAT TCTGGCGCG CTACGGGAAG TCTCCCCAC CTCATGAAGT 840
55 CAAGCTCAAG AATACACCA TTTTCTGTC GCGACCTTC TG 882

60.

(2) INFORMATION FOR SEQ ID NO: 168:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1208 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

GGGAAACTCA AAAGGATGAT GGAATGGTTG ATGGAGCCAG AGCCTAGAAG TRAAGGGATA 60
CAGAGTGAAG ATAGAGGTAT TTACGTATAT TTWAATATTA GCTTTGGAAT TACGTAGGGA 120
TTCTTAAGAA AAGATCATGA CAGGACAGCC ACATTTGGTA AAATGTCAGG GCAGCCAGTG 180
CATGGTCTCT CTGGGGCTCC TCAGTTGACG GGTTTAAATC ATTTCTGAT CCCCCTGCCC 240
TGGTTTGAGG AATGCATACA GTACGTGAAA TGCCTGTGGT ATGAGTTGCA ATGGGCAATC 300
AACCTGGGTA AATCCAAGAT TAATGATTAG TTCTAAAGAT CCAGTTGAAG TTCTAGAGTG 360
GGAATTTTCC GTCAAGCARG TCAGCACAGC TTTATGCTG TTCTCTAAT AACGATAGGT 420
AACAAATAGC TGTGKTWCA CAGCTAGGAR GATAACCAA TCTAGAGTTC TTGARTCTCA 480
TTTAATAAAT AAKTATTATG AGTACCAACT GCATATTICA GGCCTGCTAT TTGACTCTGT 540
TAAATACTGA TYCCTTAKGA CMSCCACWTC AGAWAACMIT AATCTGCTG ATCAATAAAC 600
AGCTTGACTT AGAGRGGTAA AATAGCTTGC CACAGGTWAC CCAATTAGTA GGTAAACAGCG 660
ACAGAATAAC AGTGCAGTTA AAATCTTAGA CTGGAGACTA ATTGCATAAG TTTGAATTTT 720
AGTTCGCTA TGTAAATTTG GGTGAGTACC TTAATTYACC TGAGTCTCGG TCTTTATATC 780
TGTAGAATGG AGCTAATGAT ATTACTTAAT TTGCTTTATG TGAGATTAAA TGTACTAATA 840
TATGTAAATC ACTTACAACA GCAATTGACA TATTTGACAT ACTTAATATA TTTGCTACTA 900
ATACTATTAG CAACAGCAAT CTGATTTTCC AAGTTGAAAT TCAGTGTITT CTTTTTTACT 960
TTGCCATAAT TTACAATGTT GIGCTCTGTA AACCATAAAT TTCCCTGAGG TGTGTGTCAGG 1020
TTAAAAAATA ATCACTATGG CCCCARNMA CTTGGAAAAT AGAAATGAGA CCAGCTTCAT 1080
CTATATTCTT TACTGCAAT AACTTAGAAT TGTAATAGGC TAATATGTAC TGGGACTTCC 1140
AATTTGGGAA TATGACAAA ATAATACTAT TTAGCTAAAA CATATACAGA ACTTATTTTT 1200
CCTCTGAA 1208

(2) INFORMATION FOR SEQ ID NO: 169:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1307 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

5 GGCACGAGAG AAAAGAGGTT GAGAATGTTT TCTAGCAGGC AGAATGTGCA TACATGTTTT 60
CATGARTGTC CTTTGGGTGC TGTTTCTTTT AAATCCTCTG TGCACAGGGC TCTGGCCTTT 120
ARTAAACTGT TTTTCTGTCT TACGTCATGC TGACTGGGTG CTAGGGGCTG ATTACAAAGG 180
10 GGAAGAGTTG AACAGACATC AGGGGCCGAT GAAACCAAAG GACTAGGAGT CAGGAGAACA 240
AGTCAGGGAT TAGGAGACAG CGGTTTGGTT TATTGTTATC CAGCTGGAGG ACTCCTAGGG 300
15 GCAGCAGCAG GAGGAATACC AGGGCCACGG AGGGGCAGGA GTCTCAGAGT GGAGGGCAGA 360
CTCTAACAGA TGCCAGCTGA ACGCTCGCTG GCCCTGGATG TCATACGAGT TGGGGACCAG 420
AAATCTGGGC TCAGAGAAAC CGTCCAGGA GATTTGAAGC CATGGGTTAT CTTCTAGAGT 480
20 TGATACTGAT AATATATTTT AATTTTATTT GATGTTTAAAT ACCTTCTGAA ACAGGAGGGT 540
AAGATCAGAT GGAAGGCCY TCTGTTGAAG GATCTTGGGA ACCTTGGTGG TTTTMTTTTT 600
25 TTGGTTTTTT TTTTMTTGAT CGAGCTGTGG ACATCCTTCT TAATTCGATT NTGAGGATTT 660
GTTTAACTAA AAAGTTCCCA AACACAGAAA GGGCCTCCCC ACCTGCTTTG GGGAGCTGTC 720
TGTSCTGGGA GTGCCAGGCA TCCSATGGGA CCCATCACTG CCAGTGTCTG TGCCTCCAG 780
30 AGGTCAGCCC TGTGCTGCC CTGGCTCTGT CTCCTCTGTG ACAGGGCAGA GCATTTCTGG 840
TCAGTTTCTC CATGGTGCCT CCCACCCCTT TGTAAGTGG ATGGACATGA TGAATTTCAG 900
35 TTGTCTCACC CTGATAGCCT GGGTGTGAT ATTCACTTTA CCGCACTCA GACACAGGCG 960
ACCTGAAGC AGTTCTCGGT GTGTAGAGTC CACGTGACAG TCCCCACAGC CTCCCAGAT 1020
AGCTGTGTGC CTGTGCGCTA CTGCTGTGCC ATTTTCCCAA CTNNGGCGTT TCACTAAATG 1080
40 CAGCTGATCT CTCTCTCTGT GCACTCGTGA TCCATGTTGA ACAATACATG TAGGTTCTTT 1140
TTCCACGCAA TGTAAGAACA TGATATACTG TACGTTGGAA AGCATTTACC TTATTTATAT 1200
45 ACCTGAATGT TCCTACTACA CAAATAAACA TATATTAAAT WCTAAAAAAA AAAAAAAAAA 1260
CTGGAGGGGG GCGCCGTAC CCAAATCGCC GGATAGTGAT CGTAAAC 1307

50

(2) INFORMATION FOR SEQ ID NO: 170:

(i) SEQUENCE CHARACTERISTICS:

55

- (A) LENGTH: 1624 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

	GGCAGGAGGT CGCCGCCGCG GCCGCCTGGA ATTGTGGGAG TTGTGTCTGC CACTCGGCTG	60
	CCGGAGGCGA AGGTCCCTGA CTATGGCTCC CCAGAGCCTG CCTTCATCTA GGATGGCTCC	120
5	TCTGGGCATG CTGCTTGGGC TGCTGATGGC CGCTGCTTC ACCTTCTGCC TCAGTCATCA	180
	GAACCTGAAG GAGTTTGGCC TGACCAACCC AGAGAAGAGC AGCACCAAAG AAACRGAGAG	240
10	AAAAGAAACC AAAGCCGAGG AGGAGCTGGA TGCCGAAGTC CTGGAGGTGT TCCACCCGAC	300
	GCATGAGTGG CAGGCCCTTC AGCCAGGGCA GGCTGTCCCT GCAGGATCCC ACGTACGGCT	360
	GAATCTTCAG ACTGGGAAA GAGAGGCAAA ACTCCAATAT GAGGACAAGT TCCGAAATAA	420
15	TTTGAAAGGC AAAAGGCTGG ATATCAACAC CAACACCTAC ACATCTCAGG ATCTCAAGAG	480
	TGCACTGGCA AAATTCAAGG AGGGGGCAGA GATGGAGAGT TCAAAGGAAG ACAAGGCAAG	540
20	GCAGGCTGAG GTAAAGCGGC TCTTCCGCCC CATTGAGGAA CTGAAGAAAG ACTTTGATGA	600
	GCTGAATGTT GTCATTGAGA CTGACATGCA GATCATGGTA CGGCTGATCA ACAAGTTCAA	660
	TAGTTCACG TCCAGTTTGG AAGAGAAGAT TGCTGCGCTC TTTGATCTTG AATATTATGT	720
25	CCATCAGATG GACAATGCGC AGGACCTGCT TTCTTTGGT GGTCTTCAAG TGGTGATCAA	780
	TGGGCTGAAC AGCACAGAGC CCCTCGTGAA GGAGTATGCT GCGTTTGTGC TGGGCGCTGC	840
30	CTTTTCCAGC AACCCCAAGG TCCAGGTGGA GGCCATOGAA GGGGGAGCCC TGCAGAAGCT	900
	GCTGGTCATC CTGGCCACGG AGCAGCCGCT CACTGCAAAG AAGAAGGTCC TGTTTGCACT	960
	GTGCTCCCTG CTGCGCCACT TCCCCTATGC CCAGCGGCAG TTCTTGAAGC TCGGGGGGCT	1020
35	GCAGGTCTTG AGGACCCTGG TGCAGGAGAA GGGCACGGAG GTGCTCGCCG TGCGCGTGGT	1080
	CACACTGCTC TACGACCTGG TCACGGAGAA GATGTTGCC GAGGAGGAGG CTGAGCTGAC	1140
40	CCAGGAGATG TCCCCAGAGA AGCTGCAGCA GTATCGCCAG GTACACCTCC TGCCAGGCCT	1200
	GTGGGAACAG GGCTGGTGG AGATCACGGC CCACCTCCTG GCGCTGCCCG AGCATGATGC	1260
	CGTGAGAAG GTGCTGCAGA CACTGGGCGT CCTCTGACC ACCTGCCGGG ACCGCTACCG	1320
45	TCAGGACCCC CAGCTCGGCA GGACACTGGC CAGCCTGCAG GCTGAGTACC AGGTGCTGGC	1380
	CAGCCTGGAG CTGCAGGATG GTGAGGACGA GGGCTACTTC CAGGAGCTGC TGGGCTCTGT	1440
50	CAACAGCTTG CTGAAGGAGC TGAGATGAGG CCCACACCA GGACTGGACT GGGATGCCGC	1500
	TAGTGAGGCT GAGGGGTGCC AGCGTGGGTG GGCTTCTCAG GCAGGAGGAC ATCTTGGCAG	1560
	TGCTGGCTTG GCCATTAAAT GGAAACCTGA AGGCCAAAAA AAAAAAAAAA AAAAAAAAAA	1620
55	AAAA	1624

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2003 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

10	GGCAGAGCC AGCTTGCAGG AGGAATCGGT GAGGTCTGT CCTGAGGCTG CTGTCCGGG	60
	CCGGTGGCTG COCTCAAGGT CCCTTCCCTA GCTGCTGCGG TTGCCATTGC TTCTTGCTG	120
	TTCTGGCATC AGGCACCTGG ATTGAGTTGC ACAGCTTTGC TTTATCCGGG CTTGTGTGCA	180
15	GGGCCCCGCT GGGCTCCCCA TCTGCACATC CTGAGGACAG AAAAAGCTGG GTCTTGCTGT	240
	GGCCTCCAG GCTTAGTGTT CCCTCCCTCA AAGACTGACA GCCATCGTTC TGCACGGGGC	300
20	TTTCTGCATG TGACGCCAGC TAAGCATAGT AAGAAGTCCA GCCTAGGAAG GGAAGGATTT	360
	TGGAGGTAGG TGGCTTTGGT GACACACTCA CTCTTTCTC AGCCTCCAGG ACACTATGGC	420
	CTGTTTAAAG AGACATCTTA TTTTCTAAA GGTGAATCTC CAGATGATAG GTGAACCTGA	480
25	GTTCAGATA TACCAACTTC TGCTGTATT TCCTAAATGA CAAAGATTAC CTAGCTAAGA	540
	AACTTCCTAG GGAAGTAGGG AACCTATGTG TTCCCTCAGT GTGGTTTCCT GAAGCCAGTG	600
30	ATATGGGGGT TAGGATAGGA AGAACTTTCT CGGTAATGAT AAGGAGAATC TCTTGTTTCC	660
	TCCACCTGT GTGTAAAGA TAACTGACG ATATACAGGC ACATTATGTA AACATACACA	720
	CGCAATGAAA CCGAAGCTTG GCGGCTGGG CGTGGTCTTG CAAATGCIT CCAAGCCAC	780
35	CTTAGCCTGT TCTATTACG GCAACCCCA AAGCACCTGT TAAGACTCCT GACCCCAAG	840
	TGGCATGCAG CCCCCATGCC CACCGGACC TGGTCAGCAC AGATCTTGAT GACTTCCCTT	900
40	TCTAGGCAG ACTGGGAGG TATCCAGGAA TCGGCCCTG CCCCACGGC GTTTTCATGC	960
	TGTACAGTGA CCTAAAGTTG GTAAGATGTC ATAATGGACC AGTCCATGTG ATTTCACTAT	1020
	ATACAACCTCC ACCAGACCCC TCCAACCCAT ATAACACCCC ACCCCTGTTC GCTTCCCTGA	1080
45	TGGTGATATC ATATGTAACA TTTACTCTG TTTCTGCTGA TTGTTTTTTT AATGTTTTGG	1140
	TTTGTTTTTG ACATCAGCTG TAATCATTCC TGTGCTGTGT TTTTATTAC CCTTGGTAGG	1200
50	TATTAGACTT GCACTTTTTT AAAAAAGGT TTCTGCATCG TGAAGCATT TGACCCAGAG	1260
	TGGAACCGT GGCCTATGCA GGTGGATTCC TTCAGGTCTT TCCTTTGGTT CTTTGACAT	1320
	CTTTGCTTTC ATTGCTCTCC CGTCTTGGT TCTCCAGTTC AAATTATTGC AAAGTAAAGG	1380
55	ATCTTTGAGT AGGTTCGGTC TGAAAGGTGT GGCCTTTATA TTTGATCCAC ACACGTTGGT	1440
	CTTTTAACCG TGCTGAGCAG AAAACAAAAC AGGTTAAGAA GAGCCGGGTG GCAGCTGACA	1500
60	GAGGAAGCCG CTCAAATACC TTCACAATAA ATAGTGGCAA TATATATATA GTTTAAGAAG	1560

5 GCTCTCCATT TGGCATCGTT TAATTTATAT GTTATGTTCT AAGCACAGCT CTCTTCTCCT 1620
 ATTTTCATCC TGCAAGCAAC TCAAAATATT TAAATAAAG TTTACATTGT AGTTATTTTC 1680
 AAATCTTTGC TTGATAAGTA TTAAGAAATA TTGGACTTGC TGCCGTAATT TAAAGCTCTG 1740
 TTGATTTTGT TTCCGTTTGG ATTTTGTGGG GAGGGGAGCA CTGTGTTTAT GCTGGAATAT 1800
 10 GAAGTCTGAG ACCTTCCGT GCTGGAACA CACAAGAGTT GTTGAAAGT GACAAGCAGA 1860
 CTGCGCATGT CTCTGATGCT TTGTATCATT CTTGAGCAAT CGCTCGGTCC GTGGACAATA 1920
 AACAGTATTA TCAAGAGAA AAAAAAAAAA AAAAACTCG NGGGGGGGCC CGGTACCCAA 1980
 15 TTGCCCCTAT AGTGAGCCNA TTC 2003

20

(2) INFORMATION FOR SEQ ID NO: 172:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 786 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

30 GGCACAGCGG CACGAGAAGA CTTTGGTGT TAAGAGATTA ATGTGTTAGC CAGAACAAC 60
 CATTTCTCTA CCMGTGTGTA GTCCATTTAT CTTTAAAGAT TTTCTATTGG AATAATTTTG 120
 35 AAATTACTTT CTTAGTTTTT TCAATTAAAA ACTAAGAAAA TGCTTTGTTT ATTATGAATT 180
 GCTATTCTCT TTGATTATTA TTCTTGGAGA AAGTCTATCA GACGTAATTC TTCTGATTG 240
 40 CTCTAGGCT AGAGGAAAT GTGAAAGATG ACAAATGAAA ATTTCAAAGG TTGTCAGTAG 300
 TATGACTTCT TTTATCGTTT GTCAATTATCA CAAATATATC AACATAGGAC TTTTAAAGA 360
 TATTTTGTAC ATATTGGGCC TTAGTAGGAT TTTCATGAA TTTTTTTTTT CTTTATGCC 420
 45 CAGAGAGAAA GAGCAAAGAA ATAACCAAGG GTGATGTACT CGTATTGAAG GTTTACCAA 480
 TAAGGACTGC TTTTATTATG AACTATAGTC TATATTCTAA GTAAATCAAT TTTTCTATTA 540
 TGTGTTTTTT GTTCTGTCAG GCAAGATCTC TGAACCTTAT GCAGAGGTT CTTTAAAAA 600
 50 AACAAAGTTG AATTTTTTTA TTTCTTGGAA TATTTTTTTT CATTGATTTC TCCCAAGTAG 660
 AGCAGATCA AATCTCCTTT GTACCCATG TCTTTTTTGT TTGCTATTA GCTCAGTATT 720
 55 CCGTTTCTAC ATTTTCCCTT CTAGAACCA GTCAATAAAT GACAAAAAA AAAAAAAA 780
 ACTCGA 786

60

(2) INFORMATION FOR SEQ ID NO: 173:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 1758 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

GGGACGAGCC CTGCCACCT CTGCAGCCT CTGCGCCCC GCCGAGCTGG CGGATGGAGC 60
 TGGCAACGGG GAGCGTGGC AGCCAGCGG TGGCGCGGAG GATGGATGGG GACAGCCGAG 120
 15 ATGGCGGCGG CGGCAAGGAC GCCACCGGT CGGAGGACTA CGAGAACCTG CCGACTAGCG 180
 CCTCOGTGTC CACCCACATG ACAGCAGGAG CGATGGCCGG GATCCTGGAG CACTCGGTCA 240
 20 TGTACCCGGT GGACTCGGTG AAGACACGAA TGCAGAGTTT GAGTCCAGAT CCCAAAGCCC 300
 AGTACACAAG TATCTACGGA GCCCTCAAGA AAATCATGCG GACCGAAGCT TCTGGAGGCC 360
 CTTGCGAGGC GTCAACGTCA TGATCATGGG TGCAGGGCCR GCCCATGCCA TGTATTTTGC 420
 25 CTGCTATGAA AACATGAAAA GGACTTTAAA TGACGTTTTC CACCACCAAG GAAACAGCCA 480
 CCTAGCCAAC GGTATTTTGA AAGCGTTTGT CTGGAGTTAG AAAGTTCTCT TCTTCAACAC 540
 30 GTCCCTCCCC AGGGTGTTC TCCCTGTGAC CCAGCCGCCT CGACTTGGC CCGCTTGCTC 600
 ACGAATAAAG AACTCAGAGT TGTGTGTGCA ATGCACACCC AGACACACGC ACGCACACAC 660
 ACGCGCGCGC ACACACATGC TTTTCTCTGT TCCCTCCGC TTTCTGAAGC CTGGGGAGAA 720
 35 ATCAGTGACA GAGGTGTTTT GGTTTTATGT TTATGTGGT TTTCTTTTGT ATTTTTTTTG 780
 TTTGTTTTGT TTTTAAACAT TCAAAAGCAA TTAATGATCA GACATAGGAG AAACCCTGAA 840
 40 TAGAAACAAA ACTTTTGAAT GCTGGATTCA AAAAAAAAAA AAGTTATCT GGACAGCTTC 900
 TTTGAGACTA TTTAAAACT GGTACAACAG GTCTCTACAA CGCCAAGATC TAACTAAGCT 960
 TTAAAGGTC AAGAAGTTT ATGGCTGACA AAGGACTCGC GCAACGCAGA AGGCCTTTCC 1020
 45 CACCTTAAGC TTCCGGGAT CTGGAATTT TACCCCATTT CTCTCTGTT TGTCTGAGTC 1080
 TCATCTCTCT GCAAGCAAGG GCTGAAATCA TTTTGTGTGG TTGTTTGTAG GGAGAGAGGC 1140
 50 GGGGTGGGGG GGTGCAAATC TGCCAGCAGC TCTTACGTAA GGCATGTTTT ATTGGGGAGG 1200
 GCTGAGCTTT TATTTTCTCC TCTCCAGTGG GGTGGCTTT TATGTTTCT TGTTTGGGTT 1260
 TGAATGGAA ATATGGATAG CAGCATAAAG TACTTTTATT TTGACAAAAT TCATTTTTTT 1320
 55 CAACATGGA GACATAGATT TGACCCACAA TAACTTCTCC CCTCTCTTT TACTCTGCT 1380
 CAAAAGCAT CTCTCTCCC ATTACCCAAC CTTGGTCATA AGTGTGCCTG GCTGGTTTGC 1440
 60 AGATATTTGT TCTGCTTGT AAAAATGGC CATTAGTGCA TTTATTGAGA TGATCTCTAA 1500

AGAGCTATGC CCTGACCTAC CCTGATTCT ATGACATTGG GGCCCTTCTT TTGCTGAAAC 1560
TGCCTTACGT AATGGTTTTA CTCCTTGAAA GAGATTGAC GGAATCCATT TTATGCCAAG 1620
5 TGCTGCCCTG CACTGTTTCT GCAATATGTG GTGTATGCTG TGGTGATCTT GCTGGGAATG 1680
ATTATAAGTG TGTGTGTGGT GGGGAGTGG GTATTACATG CATTGCTGAA GAGTCAAAAA 1740
10 AAAAAAAAAA AACTCGA 1758

15 (2) INFORMATION FOR SEQ ID NO: 174:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 888 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

25 CTGTTAGAAT GCCAGTTTA CTGGATGGC AACCCAACAG TGCTCCTGCC CACCTGCCCC 60
TCAATCCTCC TAGAATTCAG CCCCCAATTG CCCAGTTACC AATAAAAACT TGTACACCAG 120
CCCCAGGGAC AGTCTCAAAAT GCAAATCCAC AGAGTGASMC ACCACCTCGG GTAGAATTTG 180
30 ATGACAACAA TCCCTTTAGT GAAAGTTTTC AAGAACGGGA ACGTAAGGAA CGTTTACGAG 240
AACAGCAAGA GAGACAACGG ATCCAATCA TGCAGGAGGT AGATAGACAA AGAGCTTTGC 300
35 AGCAGAGGAT GGAATGGAG CAGCATGGTA TGGTGGGCTC TGAGATAAGT AGTAGTAGGA 360
CATCTGTGTC CCAGATTCCT TTCTACAGTT CCGACTTACC TTGTGATTTT ATGCAACCTC 420
TAGGACCCCT TCAGCAGTCT CCACAACACC AACAGCAAAT GGGGCAGGTT TTACAGCAGC 480
40 AGAATATACA ACAAGGATCA ATTAATTCAC CCTCCACCCA AACTTTTCATG CAGACTAATG 540
AGCGAGGCAG GTAGGCCCTC CTTCAATTGT TCCTGATTCA CCATCAATCC CTGTTGGAAG 600
45 CCCAAATTTT TCTTCTGTGA AGCAGGGACA TGGAAATCTT TCTGGGACCA GCTTCCAGCA 660
GTCCCCAGTG AGGCCTTCTT TTACACCTGC TTTACCAGCA GCACCTCCAG TAGCTAATAG 720
CAGTCTCCA TGTCGCCAAG ATTCTACTAT AACCCATGGA CACAGTTATC CGGGATCAAC 780
50 CCAATCGCTC ATTCAGTTGT ATTCTGATAT AATCCAGAG GAAAAAGGN AAAAAAARA 840
AMAARAARA ARAAAGGAGA TGATGATGCA GAATCCACC AAGGCTCC 888

55

(2) INFORMATION FOR SEQ ID NO: 175:

60 (i) SEQUENCE CHARACTERISTICS:

426.

(A) LENGTH: 2379 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

	GGCAGAGCTA GTGTGGACTC CATCCCCTG GAGTGGGATC ACGNCTATGA CCTCAGTCGG	60
10	GACCTGGAGT CTGCAATGTC CAGAGCTCTG CCCTCTGAGG ATGAAGAAGG TCAGGATGAC	120
	AAAGATTCTT ACCTCCGGGG AGCTGTGTSC TTATCAGGGG ACCACAGTGC CCTAGAGTCA	180
15	CAGATCCGAC AACTGGGCAA AGCCTGGATG ATAGCCGCTT TCAGATACAG CAAACCGAAA	240
	ATATCATTCT CAGCAAACT CCCACGGGGC CGGAGCTAGA CACCAGCTAC AAAGGCTACA	300
	TGAAACTGCT GGGCGAATGC AGTAGCAGTA TAGACTCCGT GAAGAGACTG GAGCACAAAC	360
20	TGAAGGAGGA AGAGGAGAGC CTTCCTGGCT TTGTTAACCT GCATAGTACC GAAACCCAAA	420
	CGGCTGGTGT GATTGACCGA TGGGAGCTTC TCCAGGCCCA GGCATTGAGC AAGGAGTTGA	480
25	GGATGAAGCA GAACCTCCAG AAGTGGCAGC AGTTTAACTC AGACTTGAAC AGCATCTGGG	540
	CCTGGCTGGG GGACACGGAG GAGGAGTTGG AACAGCTCCA GCGTCTGGAA CTCAGCACTG	600
	ACATCCAGAC CATCGAGCTC CAGATCAAAA AGCTCAAGGA GCTCCAGAAA GCTGTGGACC	660
30	ACCGCAAAGC CATCATCTCT TCCATCAATC TCTGCAGCCC TGAGTTCACC CAGGCTGACA	720
	GCAAGGAGAG CCGGGACCTG CAGGATCGCT TGTSGCAGAT GAATGGGCGC TGGGACCGAG	780
35	TGTGCTCTCT GCTGGAGGAG TGGCGGGGCC TGCTGCAGGA TGCCCTGATG CAGTGCCAGG	840
	GTTTCCATGA AATGAGCCAT GGTTTGCTTC TTATGCTGGA GAACATTGAC AGAAGGAAAA	900
	ATGAAATGTG CCCTATTGAT TCTAACCTTG ATGCAGAGAT ACTTCAGGAC CATCACAAAC	960
40	AGCTTATGCA AATAAAGCAT GAGCTGTTGG AATCCCAACT CAGAGTAGCC TCTTTGCAAG	1020
	ACATGTCCTG CCAACTACTG GTGAATGCTG AAGGAACAGA CTGTTTAGAA GCCAAAGAAA	1080
45	AAGTCCATGT TATTGGAAAT CGGCTCAAAC TTCTCTTGAA GGAGGTCAGT CGTCATATCA	1140
	AGGAACTGGA GAAGTTATTA GACGTGTCAA GTAGTCAGCA GGATTGTCTT TCCTGGTCTT	1200
	CTGCTGATGA ACTGGACACC TCAGGGTCTG TGAGTCCCAV ATCAGGAAGG AGCACCCCAA	1260
50	ACAGACAGAA AACGCCACGA GGCAAGTGTA GTCTCTCACA GCCTGGACCC TCTGTCAGCA	1320
	GTCCACATAG CAGGTCCACA AAAGGTGGCT CCGATTCTCT CCTTCTGAG CCARGGCCAG	1380
55	GTCGGTCCGG CCGCGGCTTC CTGTTTCAGAG TCCTCCGAGC AGCTCTTCCC CTTCAGCTTC	1440
	TCCTGCTCCT CCTCATCGGG CTTCCTGCC TTGTACCAAT GTCAGAGGAA GACTACAGCT	1500
	GTGCCCTCTC CAACAACCTT GCCCGTTCAT TCCACCCCAT GCTCAGATAC ACGAATGGCC	1560
60	CTCTCCACT CTGAACCTAAG CAGATGCCAT CTGCAGAAGT GCTGGTAGCA TAAGGAGGAT	1620

	CGGGTCATAA GCAATCCCAA ACTACCAACA AGAGGACCTT GATCTTGGCG AAAGCCMTCG	1680
5	GTGTGGCAGC TTTAGCCTCC TCCAGATCAC ATGTGTGCAA ATTATGGCTT CAGAGGTGGA	1740
	AGATAAACAG TGACGGGGGA ACAACAGAC AACAAGAAGG TTTGGAAGAA ATCTGGTTTG	1800
	AGACTCTGAA CCTTAGCACT AAGGAGATTG AGTAAGGACC TCCAAAGTTC CCCGGACTCA	1860
10	TGAATTCCTGG GCCCTTGGCC NATTCTGTGC ACAGCCAAGG ACTTCAGTAG ACCATCTGGG	1920
	CAGCTTCCC ATGGTGCTGC TCCAACCATC AGATAAATGA CCTCCCAAG CACCATGTCA	1980
15	GTGTGTTACA ATCTACCAAC CAACCAGTGC TGAAGAGATT TTAGAACCTT GTAACATACA	2040
	ATTTTAAAGA GCTTATATGG CAGCTTCCTT TTACCTTGT TTTCTTTGG GGCATGATGT	2100
	TTTAACCTTT GCTTTAGAAG CACAAGCTGT AAATCTAAAA GGCACCTTTT TTTAGAGGTA	2160
20	TAAAGAAAAA CTAGATGTAA TAAATAAGAT CATGGAAGGC TTTATGTGAA AAAAGTTGAA	2220
	TGTTATAGTA AAAAAAAG ATATTTATGT ATGTACAGTT TGCTAAAGCC AAGTTTGT	2280
25	TGTATTGATT TCITTGCATT TATTATAGAT ATTATAAAAT AAAAAA AAAAACAAC	2340
	TCGAGGGGGG GCCCGGTACC CAATTCGCCC TATAGTGAG	2379

30

(2) INFORMATION FOR SEQ ID NO: 176:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1348 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

40

	GCGCCTTCAC GATGCCGGCG GTCAGTGGTC CAGGTCCCTT ATTCTGCCTT CTCTCCTGC	60
	TCCTGGACCC CCACAGCCT GAGACGGGGT GTCCTCCTCT ACGCAGGTTT GAGTACAAGC	120
45	TCAGCTTCAA AGGCCCAAGG CTGGCATTGC CTGGGGCTGG AATACCCTTC TGGAGCCATC	180
	ATGGAGGTGA GGGGACGGG TGGGACCGC TATGCCCAGG GTCCCTCAA GTGCTGGAGG	240
50	GGCTGTRACT TGGTGGGGAG TGGGTCTGTC ACAGCCATCC TCTGTCCAGG GTGGGGCAAG	300
	GCTTGGGACA GTGCCAGGCA CCCCAGGACC CCTTCCAGGC TTGTCTCTG CTCCACGGCC	360
	TCAACACCCC CCACCCCTGC CCAAGCTGTT TCTCTCTGCT CTCTCTNNTT CCTTGGCCCA	420
55	GGACTTCTCT CTCTCTCTCT GCCTCTCCTT GGACCCCTGC CCTTCTCTTA CCTCTGACCT	480
	GTGAACACAC AGACACATGC TCACACACTA AGTCCCARGC ACACMSAAG GCAATGTGGA	540
60	CCAGCACAAA CCTCCACTCT CCCGGCTCCA TCCARCGGG CCTGTGGCTG GCCATGAAAA	600

428

	CTGGGGGCTA CCTGGAGGGA AGCATCCTCA TCCCAGGTGA GTGGGCACCA GCCCTTCCCT	660
	GTATGTGTGT TGTGGGTGGA AGCAGGCATG AGAGCATCTT AGCCCATAGG TTTGTATTCA	720
5	GGGACTTCCA AACCCAGACC TACAAAGAGT GTGTCTTCTA CCAGATCTTG TTCAAAAAG	780
	GGTTTGTGAT GATGGAAC TAACGATAGAG GGAGTGAGCA AGAACAATGA GGATTAGAGT	840
10	GGAGCGTGAA ATAGTCTAGG AGCATGGCTT CCAAAACATA TGCTGTGAGG TCTGTCCACC	900
	TGAGAGTTGG GCCATGGATT TAATTCTGAG CCTCTTAGCA GGCAAAGCAA AGACAGAAAG	960
	CAGATCGGCT GTGGATTCT GTCTATAAAA TGTGAGTTCT TGGCCGGGTG CGGTGGCTCA	1020
15	CGCCTGTAAT CCGCGCGCTT TGGGAGGCCA GGGCGGATGG GTCCGAGGT CAGGAGGTG	1080
	GAAACCATCC TGGCCGAAT GGTGAAGCCC TGACTCTACT AGAAGTGCAA AGATTGGCTG	1140
20	GGTGTGGTGG CGTCCGCTG TGGTCCCAGC TTCTCGGGAG GCTGAGGCGG GAGAGTTGCT	1200
	TGGGCTGGG AGGCCGAGGT TCCGGTGAGC TGAGATCTG CCATTGCACT TCAGCCTGGG	1260
	CACAGAGCCA GACTCTGGCT CAAAAA AAAA AAAA ACTCGAGGGG GGCCCGTACC	1320
25	CAATTGCGCG NATATGATCG TAAACAAT	1348

30 (2) INFORMATION FOR SEQ ID NO: 177:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1502 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

40	CTCAAATAA ATAAATAAAT AAAAATTTGT ATTCCATTGA TTTGGGTAGA CACCAGGAAT	60
	GTGCATTTCT AACAAGCTTT CCAGGCGATC CTATAGTAAG TCATCTGTGG ACTACTTTAA	120
45	GAAACTCTTC TATAGAGAAT GGAGTTGGAT TAATAATAGG TGATTTTSTA CACTGGACTG	180
	ATTCACAAGA ACCTAAACAG TAGTCCATGA AGCTGCTCAT CTGTGGTAAC TATTTGGCCC	240
	CGTCTCACTC TGAAGCAGC AGGAGATGTT GTTTACTTTG TTTCTATCC CTTTGTCTGG	300
50	AGATTAAATTT TGGAAATGAAA GTTTTCTCT CTATGCCATT CCTGGTTCTT TTCCAAAGCC	360
	TCATACAAGA GGATTAGTCA ACAATGCATG CATTACCTTT TAAAGAATG CGATATTGAT	420
55	ACCGATGCTT ACTTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT	480
	CCGCCCTTCT ATAGCATAGT TCTCTTAGG TGAATGATT CCTATAAGAT TTCTCATTAT	540
	TAAATCATGC ATTTTTCAG ATGGAATCAA TMTTTGATTT AATCTAAGCT GATATTCTCA	600
60	TTTGTTAGAA GAACAACCTA CATGCTAGAG AGAGAGGAGG AAATATACCC ACGACCACAC	660

5 AGCCAGTTAG TATCCAGTTG GTGCTGGACT CCAGCCAGGT GTCCTGCCTC ATGGTAGTTA 720
 AATGATATAT AGAAAAGGTA AATTTTAA GAAATATTTA TTAATATATT CCTATAAAAC 780
 ATTTTAAAGG TAACCACATA AAAATGGTTA ATTTTCCAT TCCAAAGTAA ATGCTAAGCA 840
 TGTTTATTAA TGAAGCAGTA CTTCTGATTA GTATATGACA TTCTGAAGTT AATTAACTC 900
 10 ATTGCACTAA ATGTGCTTC CTTGGTATAG TGGAGGATTT GAGGATTGGA ATATAGAGTA 960
 GAGTGCTTGC TTAAGCCTGG GAGCCCATCT TTATAGCTAT TTGATGTAAG AAAAGAGACA 1020
 TGGNCCATTT CTAACTATA TAAGGTGAGT GTGTCTATTC CCAGCAGATA TAAAGGAAAA 1080
 15 AGGAACTTT TTTGATTCCT ACCTTCCAG CCTCACCTAG CCATCTTCCA GCCTCAAATA 1140
 TAGAGATGTT AGTGCAAGGT CTTGGGCTCT AGGTGATCAT TTCATAAGTC CTTTACAGAT 1200
 20 AAAGAAAAAG TAGTGTGTTGT ATGTTTGTTT TTAAGTAACC CAAAAACAAA TTTATATTGT 1260
 ATTCAGCAAA ATTGGAATTC AGGTGTTTAA TTTTAGAACA TGAAGTGCCT GCTGTTTAA 1320
 GCATTGACTT GTATAAAAG AATTGCATGT CTCAGTAAG CTTATGGGTT TTCTCATTTT 1380
 25 TAGGTATATG GCTTTTAATC ATGTAAAGTG AACATTAGT TTTCTGCAT TTTATTACAG 1440
 GTTCTTTGTT GCAATAAAGA TGCTGCTGAA ATTAATTGAA AAAAAAAAAA AAAAAAACTC 1500
 30 GA 1502

35 (2) INFORMATION FOR SEQ ID NO: 178:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1637 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

45 ATTTTCTAGC CCACAAGGAC TGAAGTTCAG ATCCAAAAGT TCACTTGCTA ATTATCTTCA 60
 CAAAATGGA GAGACTTCTC TTAAGCCAGA AGATTTTGAT TTTACTGTAC TTTCTAAAAG 120
 GGGTATCAAG TCAAGATATA AAGACTGCAG CATGGCAGCC CTGACATCCC ATCTACAAAA 180
 50 CCAAAGTAAC AATTCAAACCT GGAACCTCAG GACCCGAGC AAGTGCAAAA AGGATGTGTT 240
 TATGCCGCCA AGTAGTAGTT CAGAGTTGCA GGAGAGCAGA GGACTCTCTA ACTTTACTTC 300
 55 CACTCATTTG CTTTGTAAAG AAGATGAGGG TGTGATGAT GTTAACTTCA GAAAGGTTAG 360
 AAAGCCCAA GGAAAGGTGA CTATTTTGAA AGGAATCCCA ATTAAGAAAA CTAAAAAGG 420
 ATGTAGGAAG AGCTGTTTCTG GTTTTGTTCM AAGTGATAGC AAAAGAGAAT CTGTGTGTAA 480
 60

430

TAAAGCAGAT GCTGAAAGTG AACCTGTTGC ACAAAAAAGT CAGCTTGATA GAACTGTCTG 540
 CATTTCATGAT GCTGGAGCAT GTGGTGAGAC CCTCAGTGTG ACCAGTGAAG AAAACAGCCT 600
 5 TGTAACAAAA AAAGAAAGAT CATTGAGTTC AGGATCAAAT TTTTGTCTG AACAAAAAC 660
 TTCTGGCATC ATAAACAAAT TTTGTTTCAGC CAAAGACTCA GAACACAACG AGAAGTATGA 720
 10 GGATACCTTT TTAGAATCTG AAGAAATCGG AACAAAAGTA GAAGTTGTGG AAAGGAAAGA 780
 ACATTTCAT ACTGACATTT TAAACGTGG CTCTGAAATG GACAACAAC GCTCACCAAC 840
 CAGGAAAGAC TTCCTGAAG ATACCATCCC ACGGAACACA GATAGAAAGA AGGAAAACAA 900
 15 GCCTGTATTT TTCCAGCAA TATAACAAAG AAGCTCTTAG CCCCCACGA CGTAAAGCCT 960
 TTAAGAAATG GACACCTCT CGGTACCTT TTAATCTCGT TCAAGAAACA CTTTTTCATG 1020
 ATCCATGGAA GCTTCTCATC GCTACTATAT TTCTCAATCG GACCTCAGGC AAAATGGCAA 1080
 20 TACCTGTGCT TTGAAGTTT CTGGAGAAGT ATCCTTCAGC TGAGGTAGCA AGAACCGCAG 1140
 ACTGGAGAGA TGTGTCAGAA CTCTTAAAC CTCTTGGTCT CTACGATCTT CGGGCAAAAA 1200
 25 CCATTGTCAA GTTCTCAGAT GAATACCTGA CAAAGCAGTG GAAGTATCCA ATTGAGCTTC 1260
 ATGGGATTGG TGCACCCTGA AGACCACAAA TTAAATAAAT ATCATGACTG GCTTTGGGAA 1320
 AATCATGAAA AATTAAGTCT ATCTTAAACT CTGCAGCTTT CAAGCTCATC TGTATGCAT 1380
 30 AGCTTTGCAC TTCAAAAAAG CTTAATTAA TACAACCAAC CACCTTTCCA GCCATAGAGA 1440
 TTTTAATTAG CCCAACTAGA AGCCTAGTGT GTGTGCTTTC TTAATGTGTG TGCCAATGGT 1500
 35 GGATCTTTGC TACTGAATGT GTTTGAACAT GTTTTGAGAT TTTTTTAAAA TAAATTATTA 1560
 TTTGACAACA ATCCAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1620
 40 AAAAAAAAAA AAAAAA 1637

(2) INFORMATION FOR SEQ ID NO: 179:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2911 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

GGTGGTTTTT GTTCTGCAAT AGGCGGCTTA GAGGGAGGGG CTTTTTOGCC TATACCTACT 60
 55 GTAGCTTCTC CACGTATGGA CCCTAAAGGC TACTGCTGCT ACTACGGGGC TAGACAGTTA 120
 CTGTCTCAGC TCTAGGATGT GCGTTCTTCC ACTAGAAGCT CTCTGAGGG AGGTAATTAA 180
 60 AAAACAGTGG AATGGAAAA CAGTGCTGTA GTCATCCTGT AATATGCTCC TTGTCAACAA 240

	TGTATACATT CCTGCTAGGT GCCATATTCA TTGCTTTAAG CTCAAGTCGC ATCTTACTAG	300
	TGAAGTATTC TGCCAATGAA GAAAACAAGT ATGATTATCT TCCAACACT GTGAATGTGT	360
5	GCTCAGAACT GGTGAAGCTA GTTTTCTGTG TGCTGTGTG ATCTCTGTGT ATAAAGAAAG	420
	ATCATCAAAG TAGAAATTTG AATATGCTT CCTGGAAGGA ATTCTCTGAT TTCATGAAGT	480
10	GGTCCATTCC TGCCCTTCTT TATTTCCCTG ATAACCTGAT TGCTCTCTAT GTCCCTGTCT	540
	ATCTTCAACC AGCCATGGCT GTTATCTTCT CAAATTTTAG CATTATAACA ACAGCTCTTC	600
	TATTCAGGAT AGTGCTGAAG ANGCGTCTAA ACTGGATCCA GTGGGCTTCC CTCCTGACTT	660
15	TATTTTGTG TATGTGGCC TTGACTGCCG GGACTAAAAC TTACAGCAC AACTTGGCAG	720
	GACGTGGATT TCATCAGAT GCCTTTTCA GCCCTTCCAA TTCTGCCTT CTTTTCAGAA	780
20	ATGAGTGTCC CAGAAAAGAC AATTGTACAG CAAAGGAATG GACTTTTCTT GAAGCTAAAT	840
	GGAACACCAC AGCCAGAGTT TTCAGTCACA TCCGTCTTGG CATGGGCCAT GTCTTTATTA	900
	TAGTCCAGTG TTTTATTTCT TCAATGGCTA ATATCTATAA TGAAAAGATA CTGAAGGAAG	960
25	GGAACCAGCT CACTGAARGC ATCTTCATAC AGAACAGCAA ACTCTATTTC TTTGGCATT	1020
	TGTTTAATGG GCTGACTCTG GGCCTTCAGA GGAGTAACCG TGATCAGATT AAGAACTGTG	1080
30	GATTTTTTTA TGGCCACAGT GCATTTTCAG TAGCCCTTAT TTTTGTAAT GCATTCCAGG	1140
	GCCTTTCAGT GGCTTTCATT CTGAAGTTCC TGGATAACAT GTTCCATGTC TTGATGGCCC	1200
	AGGTTACCAC TGTCATTATC ACAACAGTGT CTGTCTTGGT CTTTGACTTC AGGCCCTCCC	1260
35	TGGAATTTTT CTGGAAGCC CCATCAGTCC TTCTCTCTAT ATTTATTTAT AATGCCAGCA	1320
	AGCCTCAAGT TCCGAATAC GCACCTAGGC AAGAAAGGAT CCGAGATCTA AGTGGCAATC	1380
40	TTTGGGAGCG TTCCAGTGGG GATGGAGAAG AACTAGAAAG ACTTACCAA CCCAAGAGTG	1440
	ATGAGTCAGA TGAAGATACT TTCTAACTGG TACCCACATA GTTGCAGCT CTCITGAACC	1500
	TTATTTTCAC ATTTTCAGTG TTTGTAATAT TTATCTTTTC ACTTTGATAA ACCAGAAATG	1560
45	TTCTAAATC CTAATATTCT TTGCATATAT CTAGCTACTC CCTAAATGGT TCCATCCAAG	1620
	GCTTAGAGTA CCCAAAGGCT AAGAAATCT AAAGAACTGA TACAGGAGTA ACAATATGAA	1680
50	GAATTCATTA ATATCTCAGT ACTTGATAAA TCAGAAAGTT ATATGTGCAG ATTATTTTCC	1740
	TTGGCCTTCA AGCTTCCAAA AAACCTGTAA TAATCATGTT AGCTATAGCT TGTATATACA	1800
	CATAGAGATC AATTTGCCAA ATATTCACAA TCATGTAGTT CTAGTTTACA TGCCAAAGTC	1860
55	TTCCCTTTT AACATTATAA AAGCTAGGTT GTCTCTTGAA TTTTGAGGCC CTAGAGATAG	1920
	TCATTTTGCA AGTAAAGAGC AACGGGACCC TTTCTAAAA CGTTGGTGA AGGACCTAAA	1980
60	TACCTGGCCA TACCATAGAT TTGGGATGAT GTAGTCTGTG CTAAATATTT TGCTGAAGAA	2040

GCAGTTTCTC AGACACAACA TCTCAGAATT TTAATTTTGA GAAATTCATG GGAAATTGGA 2100
 TTTTGTGAAT AATCTTTTGA TGTMTTAAAC ATTGGTTCCC TAGTCACCAT AGTTACCACT 2160
 5 TGTATTTTAA GTCATTTAAA CAAGCCACGG TGGGGCTTTT TTCTCCTCAG TTTGAGGAGA 2220
 AAAATCTTGA TGTCATTACT CCTGAATTAT TACATTTTGG AGAATAAGAG GGCATTTTAT 2280
 10 TTTATTAGTT ACTAATTCAA GCTGTGACTA TTGTATATCT TTCCAAGAGT TGAAATGCTG 2340
 GCTTCAGAAT CATACCAGAT TGTCAGTGAA GCTGATGCCT AGGAACCTTT AAAGGGATCC 2400
 TTTCAAAGG ATCACTTAGC AAACACATGT TGACTTTTAA CTGATGTATG AATATTAATA 2460
 15 CTCTAAAAAT AGAAAGACCA GTAATATATA AGTCACTTTA CAGTGCTACT TCACACTTAA 2520
 AAGTGCATGG TATTTTTTCAT GGTATTTTGC ATGCAGCCAG TTAACCTCTG TAGATAGAGA 2580
 20 AGTCAGGTGA TAGATGATAT TAAAAATTAG CAAACAAAAG TGACTTGCTC AGGGTCATGC 2640
 AGCTGGGTGA TGATAGAAGA GTGGGCTTTA ACTGGCAGGC CTGTATGTTT ACAGACTACC 2700
 ATACTGTAAA TATGAGCTTT ATGGTGTTCAT TCTCAGAAAC TTATACATT CTGCTCTCCT 2760
 25 TTCTCCTAAG TTTCATGCAG ATGAATATAA GGTAAATATC TATTATATAA TTCATTTGTG 2820
 ATATCCACAA TAATATGACT GGCAAGAATT GGTGGAAATT TGTAAATAAA ATAATTATTA 2880
 30 AACCTAAAAA AAAAAAAAAA AAAAATCTGA G 2911

35 (2) INFORMATION FOR SEQ ID NO: 180:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 519 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

45 GGACAGAGCC CCAGGCCAGC CAGGGCCAGG CCTACTTTGG CCACCTTAA ATTAGAATGT 60
 GGGGTCAGGG GTCACAGAAA AGCCATTCT CTGACCTAGT GTTTGGCGTC CGGGAACTCT 120
 GTGCCCAACC TTCAGACCTT GGCAGTCTCT ACTGAGGCCA TTGGCCAGAG GCGGCCATC 180
 50 CCCCAGARACC CCGGGAGACC GCCTGTTGCC ACGTCCACAC CTGOCACACC CTCTGCCGGG 240
 CCCCAGCCCC TCCCAACCGG GACCGTGCTG GTCCCTGGGG GTCCCTGCCC ACCTTGCCCT 300
 55 GGGGAGGCAT GGGCCCTCTT CCTCCACCC TGCCGGCGGT CACTCACCTC TTGCTTCTGG 360
 TCCCCAGGC CTAGCCCTTG GAAGGAGACA GGAGTCTAGG GAGGCTGAAG CCCACTCCCG 420
 60 GGGAGGCCCG TGCTCTCTCA GCCCCAGGGA CAGCAAGGAA AAGAGAAGAG AGCAGAGCAT 480

TTCATGGCTC TAATAAAAAA AAAAAAAAAA AAAACTCGA

519

5

(2) INFORMATION FOR SEQ ID NO: 181:

(i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 968 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

15

TCCCTTGGG	GCCGAAAAA	GCGGGTTGG	CCTGNCATT	GGTINCCAT	GCCGCCCGCC	60
CATGCCCCAG	TACTAGCCTG	CAGTCCCAAT	GTAGCCCCCTC	CCTCYTCCMA	GAGCCCYTCM	120
AACCGCCCCG	STCANTTGTG	ATTTCAGGAG	GATTTGATGA	AGATGTTAAA	GCGAAAGTGG	180
AGAACCTTCT	CGGGATTTC	AGCCTGGAAA	AAACGGACCC	TGTTAGGCAA	GCACCTGCA	240
GCCCTCCCTG	TCCCTTCTT	CCCTCCCT	TCYCCCGCC	GTGGAGACAG	CTGTTYTCAG	300
25	CAGGGCTCTC	CGCAGGAGG	GGCCCGGCTC	CTTCCCTGGC	AGCAACATCC	360
	CACACAAGTC	AGCCTCCATC	TGCGCAGCTC	TGTGGATGCG	CTGCTGGAGG	420
30	TGTCACTGGC	TGTTTCAGCC	CCTACCACCG	CCAGCGGAAG	CTCATCCACC	480
	TCAGCACATC	CAGCCCGCAG	CGCTCAGCCT	CCTGGCACAG	TGGAGCACCC	540
	GCTGGAGGCT	GCCCTGCAGC	TGGCTTTCTA	CCCGGATGCC	GTGGAGGAGT	600
35	AAACGTGCAC	CCCAGCCTGC	AGCGGCTGCA	ARCTCTGCTG	CAGGACCTCA	660
	TGCCCCCCCC	CTGCCACCCA	CCAGCCCTGG	CAGGGACGTT	GCTCAGGACC	720
40	GAGCTCATGC	CAGGGGGCTC	CTGCTGGAGG	CTGGGGGGGC	TCTGCWYTKY	780
	GGGCAATACG	GCCACGTGG	GCGTCGTGCC	CTCTGGCCCA	GCAGTGTCTT	840
	AGTTCTCTAG	GGCCCTGGGC	AGCCCTGGG	GGAGAGACTA	GAAAACACAG	900
45	CACAGGGAGA	CCCGCTTTGT	GATCTGCATG	TGTGACACTG	ATTCTTTGGA	960
	GGAAGCTG					968

50

(2) INFORMATION FOR SEQ ID NO: 182:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1128 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

	TGTAAAAGTT ATCAGTAATC CTAATTCITT TCCTGGGTTT TCCTTTTGTC ACTTATTAAT	60
5	CAGTTTTTGA AAGGACGAAT GAATTTAGAG ATGTACTCTG GAGCAGTATC ATGTTAAACC	120
	AGGGGTATAT TAGAAAAATC ATCCTCATAA TCATTCTGGG AAGTTTTTCC TCCCCAAAAA	180
	AAGCCATCCT GATGGGTTTT CAAAACCAGA AAAAAGCTCT TAATGAGGAA CAGACCACTG	240
10	GAGTACCCAT GAGCATCTCA GGAAACTGA GACCCTCGAG AAGCCTTGAT TTCGTGCAAC	300
	CCCCAAGGTT TCAGAGCCAG CAGCCAGTG CTGTGGTTGA CAGACGTGGT TTTKTGGRGA	360
15	AAGCAGCCAG AGGCCAGGAA TTTCAGAGT CGTGAGTCAC GRTYTCCAC CCAAGATTAG	420
	AGCAMAGATT AGCCATACTG AGATTGGTA AAATCATCTT GTCTAAGCAA TGGAGGTGTG	480
20	TGCAMAOGTG CAGTGCCTGT TCACAGGGGA TGCAGGCAGA TCSYGGGTTT AGGATGGGGR	540
	AGGCCACCGC ACCCCCYTTC AYTGCTCTGC ACCTGCTCCC TCACGTGGAC ACTGTCCACA	600
	ACTGTGGCTC TCACAGGACA GTTGCCCAAG GAGCTCATAT CTTATTGGAG ATAGGGGGTC	660
25	GTACAGGTGA CATTCATGAG CAGTGTGAGC CGGGTGACAT GGGGGTGTC ACCCAGCATC	720
	TGTCCAGGAG CTCCTCTGTC AGCGGCTCTG GCAGGTGGCC TGAGGCTCCT TTTTGAGAGA	780
	GAAGTGTTTG GCCTTCCTGT CTCCTCTCCT CTGATCTGTT CTTTCTTGGA ACACCACCCA	840
30	AGAAGTCAC CTCCTCCATC AGATTGTGAG CTCCTGGAGG GCAGGAGCTG TGTCTTCTA	900
	TTCATCTTCC TATCCCCAGA ACCTTGACA GATCCTGGAA TGTGGTAGGT GCTCAGTAAA	960
35	TGTGTGTGA ATAAATGAAT GAATGAATGA ACAAATGAAT GAATTTGCTT ACTTCAAGGC	1020
	AAAAGAACCA TGAACTGTA TTTGAGTTT CTATGTTATA GCAGTCAGCA AATCCTATTA	1080
40	AATACTTTGT GTTCCAAGC AAAAAAAAAA AAAAAAAAAA AAACTCGA	1128

(2) INFORMATION FOR SEQ ID NO: 183:

45

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2276 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

	CCGCGGGGTC TGACCTCATG GCGTAGAGCC TAGCAACAGC GCAGGCTCCC AGCCGAGTCC	60
55	GTTATGGCCG CTGCCGTCCC GAAGAGGATG AGGGGGCCAG CACAAGCGAA ACTGCTGCCC	120
	GGGTGGCCA TCCAAGCCCT TGTGGGGTTG GCGGGGCCG TGGTCTTGGC GCTCCTGCTT	180
60	GTTCCGCGG CTCTATCCAG TGTTGTATCA CGGACTGATT CACCGAGCCC AACCGTACTC	240

	AACTCACATA TTTCTACCCC AAATGTGAAT GCTTTAACAC ATGAAAACCA AACCAAACCT	300
	TCTATTTCCC AAATCAGCAC CACCTCCCT CCCACGACGA GTACCAAGAA AAGTGGAGGA	360
5	GCATCTGTGG TCCCTCATCC CTCGCCTACT CCTCTGTCTC AAGAGGAAGC TGATAACAAT	420
	GAAGATCCTA GTATAGAGGA GGAGGATCTT CTCATGCTGA ACAGTTCTCC ATCCACAGCC	480
10	AAAGACACTC TAGACAATGG CGATTATGGA GAACCAGACT ATGACTGGAC CACGGGCCCC	540
	AGGGACGACG ACGAGTCTGA TGACACCTTG GAAGAAAACA GGGGTTACAT GGAAATGAA	600
15	CAGTCAGTGA AATCTTTTAA GATGCCATCC TCAAATATAG AAGAGGAAGA CAGCCATTTT	660
	TTTTTTCATC TTATTATTTT TGCTTTTTCG ATTGCTGTGG TTTACATTAC ATATCACAAC	720
	AAAAGGAAGA TTTTCTTCTT GGTTCAAAGC AGGAAATGGC GTGATGGCCT TTGTTCCAAA	780
20	ACAGTGGAAAT ACCATCGCCT AGATCAGAAT GTTAATGAGG CAATGCCTTC TTTGAAGATT	840
	ACCAATGATT ATATTTTTTA AAGCACTGTG ATTTGAATTT GCTTATGTAA TTTTATTTGC	900
	TTGACTTTTT ATATGATATT GTGCAAATGT TTGCCATAGG CAATTGGTAC TTAAATGAGA	960
25	GGTGAGTCTC TCTTTTGCCT TGGTGCTTTG GAAATTAAAT GTCACAAACG AGTATATAAT	1020
	TTTTTATCTG TACTTTTAGA GCTGAGTTTA ATCAGGTGTC CAAATGTGA GTTAAACATT	1080
30	ACCTTATATT TACACTGTTA GTTTTATTG TTTTATGATT ATTATGCTTC TTCTGGAAGT	1140
	ATTAGTGATG CTACTTTTAA AAGATCCCAA ACTTGTAAC TAAATCTGAC ATATCTGTTA	1200
35	CTGCTGACTC ACATTCATTC TCCGCCATTC AAATACTATT TTTTATCCAC ATTTTTTTTT	1260
	GTTCCTCAAAC TGTAATGTAC AAGGATATGT GTGATAATGC TTTGGATTG AGTAATATTT	1320
	TTTTTTCTTC CAAGAAACT GCTTTGATA TTTTATAGATA ATTAAACAT AATTAGGAT	1380
40	AATGATATTG CTCAATCTGA CCACAATTTT AGGTAAAACA TTAAATGIGT CAGAAATCTT	1440
	GGCAACAGAG ACTCTGCAGC TTGCACTGGA CATAGATAAA ATGTTACAGA GATACTATTT	1500
	TTTTGGTTGG AATTACTATA TTAAATTTAG AAGCAGAAAC TGGTAAAATG TTAAATACAT	1560
45	GTACAATTGC TTTTAGTTAG CAATTGATTG TAGCATGGGT TCCTCCAAGG TTTCAAGCAA	1620
	TGGGCAGAGT TTAAATTTAT ATCAGATTCT TTTACTTCGT TTATTATTTT ACAGTAAATT	1680
50	TGAATAAATC TTAGGGGTCA TTATCACTTA AATAATACTG TACCTAGGTC TTTCAAATTA	1740
	AAATTATACC TGAATGAAGT TGTTTGTATA CATAAAGGAT ATTTGTGTAC AATTACCTTT	1800
55	TTTCCCCAC ACTGTMTTC TTGTMTTTG TTTTATATGG CAACTGGAAA GTATTTACTA	1860
	TGGGATTCAT TTAGTCTGCT CTTTCTATCA TAAAGAATTG ATCAATATGT AAATATGTGA	1920
	TTTGAACCAT GGTGACTTA CAAGTGTAC TACAGCTTTT TAGAAAACAT AGCCCTAATA	1980
60	TATGTTAAGC AGGACCCGGG TGAGCCAGTG GGCTTGCCT TTATGTAGAG CTGGAAGAAG	2040

5 GCGGTCCATC CTGTCTCTTG GCGGACAGT GTACTTTCCT AATAGGGAAG GGAAGCACAA 2100
TGGAATACC CCTGAACCGT TTTATTGCAG TAATTTTTTT CATATCTGAA ACTATTATTT 2160
AATATTTTGA ATAAGATTIT AAAAAATAAA TGGCAAAGAT ATAAATCTAA AAAAAAAAAA 2220
AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA 2276

10

(2) INFORMATION FOR SEQ ID NO: 184:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2500 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

TCCAAGCTAC GCCACTCGGG CTGGGGCGTT GGGAGCGGGA GTGCAGAGCG TGGTCGTGGC 60
25 GCGGCGCGTG AGAAGAGCGA GCGKAGGAG GGGTGCCAT GGCCGGGCAG CAGTTCCAGT 120
ACGATGACAG TGGGAACACC TTCTTCTACT TCCTCACCTC CTTCGTGGGG CTCATCGTGA 180
TCCCGGCGAC ATACTACCTC TGGCCCCGAG ATCAGAATGC CGAGCAAATT CGATTAAAGA 240
30 ATATCAGAAA AGTATATGGA AGGTGTATGT GGTACGTTTA CGGTTATTAA AACCCAGCC 300
AAATATTATT CCTACAGTAA AGAAAATAGT TCTGCTGCA GGATGGGCAT TGTTCATTAT 360
35 CCTTGCATAT AAAGTTTCCA AAACAGACCG AGAATACCAA GAATACAATC CTTATGAAGT 420
ATTAAATTG GATCCTGGAG CCACAGTAGC AGAAATTAAA AAACAATATC GTTGTCTGTC 480
ACTTAAATAT CATCCAGATA AAGGAGGTGA TGAGGTTATG TTCATGAGGA TAGCAAAAGC 540
40 TTATGCTGCT TTAACGGATG AAGAGTCCCG GAAAAATTGG GAAGAATTG GAAATCCAGA 600
TGGGCCTCAA GCCACAAGCT TTGGAATTGC CCTGCCAGCT TGGATAGTTG ACCAGAAAAA 660
45 TTCAATTCTG GTTTTACTTG TATATGGATT GGCATTTATG GTTATCCTTC CAGTTGTGT 720
GGGCTCTTGG TGGTATCGCT CAATACGCTA TAGTGGAGAC CAGATTCTAA TACGSACAAC 780
ACAGATTTAT ACATACTTTG TTTATAAAAC CCGAAATATG GATATGAAAC GTCTTATCAT 840
50 GGTPTTGGST GGAGCTTCTG AATTGATCC TCAGTATAAT AAAGATGCCA CAAGCAGACC 900
AACGGATAAT ATTCTAATAC CACAGCTAAT CAGAGAAATT GGCAGCATTA ATTTAAAGAA 960
55 GAATGAGCCT CCACCTACCT GCCATATAG CCTGAAGGCC AGAGTTCTTT TACTGTCTCA 1020
TCTTGCTAGA ATGAAATTC CTGAGACCCT TGAAGAAGAT CAGCAATTCA TGCTAAAAA 1080
GTGTCTGCC CTACTTCAAG AAATGGTTAA TGTAATCTGC CAACTAATAG TAATGGCCCG 1140
60

437

GAACCGTGAA GAAAGGGAGT TTCGTGCTCC AACTTTGGCA TCCCTAGAAA ACTGCATGAA 1200
 GCTTTCTCAG ATGGCCGTTT AGGGACTTCA GCAATTTAAG TCTCCCTTC TGCAGCTCCC 1260
 5 TCATATTGAA GAGGACAATC TTAGACGGGT TTCTAATCAT AAGAAGTATA AAATTAAAAC 1320
 TATCCAGGAT TTGGTGAGTT TAAAAGAATC AGATCGTCAC ACTCTACTGC ACTTCCTTGA 1380
 AGATGAAAAA TATGAAGAGG TTATGGCTGT CCTTGGGAGT TTTCCATATG TGACCATGGA 1440
 10 TATAAAATCA CAGGTGTTAG ATGATGAAGA TAGCAACAAC ATCACAGTAG GATCCTTAGT 1500
 TACAGTGTG GTTAAGTTGA CAAGGCAAAC AATGGCTGAA GTATTTGAAA AGGAGCAGTC 1560
 15 CATCTGTGCT GCAGAGGAAC AGCCAGCAGA AGATGGGCAG GGTGAAACTA ACAAGAACAG 1620
 GACAAAAGGA GGATGGCAAC AGAAGAGTAA AGGACCCAAG AAAACTGCTA AATCAAAAAA 1680
 AAAGAAACCT TTAAAAAAA AACCTACACC TGTGCTATTA CCACAGTCAA AGCAACAGAA 1740
 20 ACAAAGCAG GCAATGGAG TCGTTGGGAA TGAAGCTGCA GTAAAGGAAG ATGAAGAAGA 1800
 AGTTTCAGAT AAGGGCAGTG ATTCTGAAGA AGAAGAAACC AATAGAGATT CCCAAAGTGA 1860
 25 GAAAGATGAT GGTAGTGACA GAGACTCTGA TAGAGAGCAA GATGAAAAAC AAAACAAAGA 1920
 TGATGAAGCA GAGTGGCAAG AATTACAACA AAGCATACAG CGAAAAGAGA GAGCTCTATT 1980
 GGAAACCAA TCAAAAATAA CACATCCTGT GTATAGCCTT TACTTTCCTG AGGAAAAACA 2040
 30 AGAATGGTGG TGGCTTTACA TTGCAGATAG GAAGGAGCAG ACATTAATAT CCATGCCATA 2100
 TCATGTGTGT ACGCTGAAAG ATACAGAGGA GGTAGAGCTG AAGTTTCCTG CACCAGGCAA 2160
 35 GCCTGGAAAT TATCAGTATA CTGTGTTTCT GAGATCAGAC TCCTATATGG GTTTGGATCA 2220
 GATTAAACCA TTGGAAGTTK GGAAGTTCAT GAGGCTGAAG CCTGTGCCAG AAAATCACCC 2280
 ACAGTGGGAT ACAGCAATAG AGGGGGATGA AGACCAGGAG GACAGTGAGG GCTTTGAAGA 2340
 40 TAGCTTTGAG GGAGGAAGAG GGAGGGAGGA AGGAAGGTGG TGGACTTAAG GCAGTTACTC 2400
 TGAATGGGA CCCACAGTGT TTTGCACCAT ATTTTGCAA TTTTMTTTC CCGTTTITNG 2460
 45 GAAGTGTTTT CCNTNAANCC CAGGAACCAT TACAGAACCG 2500

50 (2) INFORMATION FOR SEQ ID NO: 185:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1337 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

60 CTTCCGGTTC TCCGGGCAGC TGCCACTGCT GTAGCTTCTG CCACCTGCCA CGACCGGGCC

60

TCTCCCTGGC GTTTGGTCAC CTCTGCTTCA TTCTCCACCG CGCTATGGT CCTCTTGGA 120
GCCAGCGTGG CGGGCCTGGC GGCTCCCGGG TGGTGAGAGA GCGGTCCGGG AACGATGAAG 180
5 GCCTCGCAGT GCTGCTGCTG TCTCAGCCAC CTCTTGGCTT CCGTCTCTCT CCTGCTGTG 240
CTGCTGAAC TAAGCGGGYC CTGGMAGTC CTGCTGCAGG CAGCCGAGGC CGCGCCAGGT 300
10 CTTGGGCCTC CTGACCCTAG ACCACGGACA TTACCGCCGC TGCCACGGG CCTACCCCT 360
GCCCAGCAGC CGGGCCGTGG TCTGGCTGAA GCTGCGGGGC CGCGGGGCTC CGAGGGAGGC 420
AATGGCAGCA ACCCTGTGGC CGGCCTTGAG ACGGACGATC ACGGAGGGAA GGCCGGGGAA 480
15 GGCTCGGTGG GTGGCGGCCT TGCTGTGAGC CCCAACCCCTG GCGACAAGCC CATGACCCAG 540
CGGGCCCTGA CCGTGTGAT GGTGGTGAGC GGCGCGGTGC TGGTGTACTT CGTGGTCAGG 600
20 ACGGTCAGGA TGAGAAGAAG AAACCGAAAG ACTAGGAGAT ATGGAGTTTT GGACACTAAC 660
ATAGAAAATA TGAATTGAC ACCTTTAGAA CAGGATGATG AGGATGATGA CAACACGTTG 720
TTTGATGCCA ATCATCTCG AAGATAAGAA TGTGCCTTTT GATGAAAGAA CTTTATCTTT 780
25 CTACAATGAA GAGTGAATT TCTATGTTTA AGGAATAAGA AGCCACTATA TCAATGTTGG 840
GGGGGTATTT AAGTTACATA TATTTTAACA ACCTTTAATT TGCTGTGCA ATAAATACCG 900
30 TATCCTTTTA TTATATCTTT ATATGTATAG AAGTACTCTR TTAATGGCT CAGAGATGTT 960
GGGATAAAG TATACTGTAA TAATTATCT GTTTGAAAAT TACTATAAAA CGGTGTTTTT 1020
TGATCGGTTT TTGTTTCTG CTTACCATAT GATTGTAAAT TGTTTATGT ATTAATCAGT 1080
35 TAATGCTAAT TATTTTGTCT GATGTCATAT GTTAAAGAGC TATAAATTCC AACAACCAAC 1140
TGGTGTGTAA AAATAATTAA AAATTTCCTT TACTGAAAGG TATTTCCCAT TTTTGTGGG 1200
40 AAAAGAAGCC AAATTATTA CTTGTGTG GGGTTTTTAA AATATTAAGA AATGTCTAAG 1260
TTATGTGTTG CAAAACAATA AATATGATTT TAAATTCTCT TAAAAA AAAAACC 1320
CCGGGGGGGG GCCCGN 1337
45

(2) INFORMATION FOR SEQ ID NO: 186:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 941 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

GGCAGGAGCC TGGACGCAGC AGCCACCGCC GCGTCCCTCT CTCCAGGAGG CTGCCGGCTT

60

60

AGGACCCCCA GCTCCGACAT GTCGCCCTCT GGTCGCCTGT GTCTTCTCAC CATCGTTGGC 120
 CTGATTCTCC CCACCAGAGG ACAGACGTTG AAAGATACCA CGTCCAGTTC TTCAGCAGAC 180
 5 TCAACTATCA TGGACATTCA GGTCCCAGACA CGAGCCCCAG ATGCAGTCTA CACAGAACTC 240
 CAGCCCCACCT CTCCAACCCC AACCTGGCCT GCTGATGAAA CACCACAACC CCAGACCCAG 300
 ACCCAGCAAC TGGAAAGAAC GGATGGGCCT CTAGTGACAG ATCCAGAGAC ACACAAGAGC 360
 10 ACCAAAGCAG CTCATCCAC TGATGACACC ACGACGCTCT CTGAGAGACC ATCCCCAAGC 420
 ACAGACGTCC AGACAGACCC CCAGACCCCTC AAGCCATCTG GTTTTCATGA GGATGACCCC 480
 15 TTCTTCTATG ATGAACACAC CCTCCGAAA CGGGGGCTGT TGGTCGCAGC TGTGCTGTTT 540
 ATCACAGGCA TCATCATCCT CACCACTGGC AAGTGCAGGC AGCTGTCCCG GTTATGCCCG 600
 AATCATGTCA GGTGAGTCCA TCAGAAACAG GAGCTGACAA CCYGCTGGGC ACCCGAAGAC 660
 20 CAAGCCCCCT GCCAGCTCAC CGTGCCAGC CTCCTGCATC CCTCGAAGA GCCTGGCCAG 720
 AGAGGGAAGA CACAGATGAT GAAGCTGGAG CCAGGGCTGC CGGTCCGAGT CTCCTACCTC 780
 25 CCCCACCCCT GCCCGCCCT GAAGGCTACC TGGCGCCTTG GGGGCTGTCC CTCAGTTAT 840
 CTCCTCTGYT AAGACAAAA GTAAAGCACT GTGGTCTTTG CAAAAA AAAA AAAAAA 900
 AAAAAA AAAA AAAAAA AAAAAA AAAAACTCG A 941
 30

(2) INFORMATION FOR SEQ ID NO: 187:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 654 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

GAATTCGGCA CGAGGCAGCT TGTGCTTTAA AGGAGGTGTT CAAAGCATGT CTGAGCAGAG 60
 45 ACTTTTGGGC TCTGTTTAA TTAATACTTT AAAATAATTC ATATTTAAAA TATCARATGT 120
 TTCCATAAAG AGGAGGATGT TTAATGCCT CCAGACTACA TTCCTTTTTA TTSCITGATT 180
 50 TTACCTGGGA GTCCAAAGTT CAATCCCAT AAAGCAAGCG TTTTATTTGT CACTTTCAAT 240
 ATACATCCGA TTGCCATGCT TAAGATGCAA TATGGGCTGC GGAATAGGT TAACCCACAG 300
 GCTCCAGGG CCCAGTGTAG AAGGTGAGAG ATTCTGTGTA AATGATTCAA ATAAAAGGAA 360
 55 GACCCCTGGC GGGTGCCGTA RCTCAGCCT GTAATCCAG CACTTTGGGA GGCCGAAGCG 420
 AGTGATGAC GAGGTTAGGA GTTGAGACC AGCCTGGCCA ACATCGTGAA ACCCCGTCTC 480
 60 TACTAAAAAT AAAAAAATTA GCGGGCATG GTGCAGGCA CCTGTAATCC TAGCTAGTTG 540

GGAGGCTGAG GCAGGAGAAT CGTTTGAATC TGGGAGTTGG AGGTTGTCAG TGAGCTGAGA 600
TCGCGCCACA GCACTCCAGC CTGGGTGACA GGGTGAGACT CTGTCTCAAA NAGA 654

5

(2) INFORMATION FOR SEQ ID NO: 188:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1848 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

15

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

20 GAAACTGGAC CGGAGAACCG GAGCGAAGCG AAGCGGAAGC CCGGAATGAG GCCGGACTGG 60
AAAGCCGGAG CGGGGCCAGG CGGGCCTCCC CAAAAGCCTG CCCCTTCATC CCAGCGGAAA 120
CCGCGGCCCC GGCCGAGCGC GCGCGCCGCT GCGATTGCAG TCGCGGCGGC GGAGGAAGAG 180
25 AGACGGCTCC GGCAGCGGAA CGGCCTGAGG CTGGAGGAGG ACAAACCGGC CGTGGAGCGG 240
TGCTTGAGG AGCTGGTCTT CGGCGACGTC GAGAACGACG AGGACGCGTT GCTGCGGCGT 300
CTGCGAGGCC CGAGGGTCA AGAACATGAA GACTCGGGTG ACTCAGAAGT GGAGAATGAA 360
30 GCAAAAGGTA ATTTTCCACC TCAAAGAAG CCAGTTTGGG TGGATGAAGA AGATGAAGAT 420
GAGGAAATGG TTGACATGAT GAACAATCGG TTTGGAAGG ATATGATGAA AAATGCTAGT 480
35 GAAAGTAAAC TTTGAAAGA CAACCTTAAA AAGAGACTTA AAGAAGAATT CCAACATGCC 540
ATGGGAGGAG TACCTGCCGT GGCAGAGACT ACTAAGCGGA AAACATCTTC AGATGATGAA 600
AGTGAAGAGG ATGAAGATGA TTTGTTGCAA AGGACTGGGA ATTTTCATATC CACATCAACT 660
40 TCTCTTCCAA GAGGCATCTT GAAGATGAAG AACTGCCAGC ATGCGAATGC TGAACGTCCT 720
ACTGTTGCTC GGATCTCCAT CTGTGCAGTT CCATCCCGGT GCACAGATTG TGATGGTTGC 780
45 TGGGATTAGA TAATGCTGTA TCACTATTTT AGGTTGATGG GAAACAAAT CCTAAAATTC 840
AGAGCATCTA TTTGAAAGG TTTCCAATCT TTAAGGCTTG TTTAGTGCT AATGGGGAAG 900
AAGTTTTAGC CACGAGTACC CACAGCAAGG TTCTTTATGT CTATGACATG CTGGCTGGAA 960
50 AGTTAATTC TGTGCATCAA GTGAGAGGTT TGAAGAGAA GATAGTGAGG AGCTTTGAAG 1020
TCTCCCAGA TGGTCTCTC TTGCTCATAA ATGGCATTGC TGGATATTG CATTGCTAG 1080
55 CAATGAAGAC CAAAGAACTG ATTGGAAGCA TGAAAATTAA TGAAGGGTT GCAGCATCCA 1140
CATTCCTTC AGATAGTAAG AAAGTATACG CCTCTTGGG GATGGAGAA GTTTATGTTT 1200
GGGATGTGAA CTCAAGGAAG TGCCTTAACA GATTTGTTGA TGAAGGCAGT TTATATGGAT 1260
60

TAAGCATTCG CACATCTAGG AATGGACAGT ATGTTGCTTG TGGTCTAAT TGTGGAGTGG 1320
 TAAATATATA CAATCAAGAT TCTTGCTCTC AAGAAACAAA CCCAAAGCCA ATAAAAGCTA 1380
 5 TAATGAACTT GGTACAGGT GTTACTTCTC TGACCTTCAA TCCTACTACA GAAATCTTGG 1440
 CAATTGCTTC AGAAAAATG AAAGAAGCAG TCAGATTGGT TCATCTTCCT TCCTGTACAG 1500
 TATTTTCAAA CTTCCAGTC ATTAAAAATA AGAATATTTC TCATGTCAT ACCATGGATT 1560
 10 TTTCTCCGAG AAGTGGATAC TTTCCTTGG GGAATGAAAA GGGCAAGGCC CTGATGTATA 1620
 GGTGCACCA TTAATCAGAC TTCTAAAGAG ACTATTTGAA GTCCAGTTGA GTCACAAGAG 1680
 15 AAGCCTGTCT TGATATATCA TCTCAGAAAC TTCTCTGAAT ATGTGATAAT ATATGGAAAA 1740
 TGATTTATAG ATCCAGCTGT GCTTAAGAGC CAGTAATGTC TTAATAAACA TGTGGCAGCT 1800
 TTTGTTTGA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AACTCGA 1848
 20

(2) INFORMATION FOR SEQ ID NO: 189:

25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1146 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

30

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

AAAAAAACC CAGGGGAACN TTGGGGGCCG CTTNNNTTC CCCCTCCAGG CCATTGGGGA 60
 35 ATTCTTCAAG TTAATCTGC TTGCTCTTG GCCAACAGGG CTTGTAGGGG GGAGAGACCC 120
 AGGATCATCA AGGGGTTGGA GTGCAAGCCT CACTCCAGC CCTGGCAGGC AGCCCTGTTC 180
 40 GAGAAGACGC GGCTACTCTG TGGGGGACG CTCATCGCCC CCAGATGGCT CCTGACAGCA 240
 GCCCACTGCC TCAAGCCCCG CTACATAGTT CACCTGGGGC AGCACAACCT CCAGAAGGAG 300
 GAGGGCTGTG AGCAGACCCG GACAGCCACT GAGTCCTTCC CCCACCCCGG CTTCAACAAC 360
 45 AGCCTCCCCA ACAAAGACCA CCGCAATGAC ATCATGCTGG TGAAGATGGC ATCGCCAGTC 420
 TCCATCACCT GGGCTGTGCG ACCCCTCACC CTCTCTCAC GCTGTGTCAC TGCTGGCACC 480
 50 AGCTGYCTCA TTTCGGCTG GGCAGMACG TCCAGCCCCC AGTTACGCCT GCCTCACACC 540
 TTGSGATGG CCAACATCAC CATCATTGAG CACCAGAAGT GTGAGAAGC CTACCCCGGC 600
 AACATCACAG ACACCATGGT GTGTGCCAGC GTGCAGGAAG GGGCAAGGA CTCCTGCCAG 660
 55 GGTGACTCCG GGGGCCCTCT GGTCTGTAAC CAGTCTCTTC AAGGCATTAT CTCCTGGGGC 720
 CAGGATCCGT GTGCGATCAC CCGAAAGCCT GGTGTCTACA CGAAAGTCTG CAAATATGTG 780
 60 GACTGGATCC AGGAGACCAT GAAGAACAAT TAGACTGGAC CCACCCACCA CAGCCCATCA 840

COCTCCATTT CCACTTGGTG TTTGGTTCCT GTTCACTCTG TTAATAAGAA ACCCTAAGCC 900
AAGACCTCT ACGAACATTC TTTGGGCCTC CTGGACTACA GGAGATGCTG TCACTTAATA 960
5 ATCAACCTGG GGTTCGAAAT CAGTGAGACC TGGATTCAAA TTCTGCCTTG AAATATTGTG 1020
ACTCTGGGAA TGACAACACC TGGTTTGTTC TCTGTTGTAT CCCCAGCCCC AAAGACAGCT 1080
10 CCTGGCCATA TATCAAGGIT TCAATAAATA TTTGCTAAAT GAAAAAARAA AAAAAAAAAA 1140
ACTCGA 1146

15

(2) INFORMATION FOR SEQ ID NO: 190:

(i) SEQUENCE CHARACTERISTICS:
20 (A) LENGTH: 906 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

ACTCCCTCAC CCAGGTCCCA GCCCTGGGAA CCACCTACCG TGAGCCCTTT TGCAGATATA 60
GACTCATTTC ATCCTCAGAT GGTCCCTCAA GGTAGGTACT TTAGTCCCAT TTTAGAGATG 120
30 AGACGATTGA GGCCAGAGGG GTGNGTAAC TTGCCTGGGG GCTCAGGAGC ACAAAGGAG 180
CCGAGGCAGG ATCTGACCCT TGTCTCTGG CCTCACTGCC CTCACTTTGC CATGACCCGA 240
35 AGTTATGTCC CTACAAAGCA ATGCATGGTC CAAGGYTCTT TTTATTGTAT TTTTATTTT 300
AAGGGTCTG TTCAAACTG GTGTGAGCTC TGAGGAGTCC TGAACCCTGG GTGCAGCATC 360
CTAGCATCCT GGGAGTCTT TTCTGCCAC ACTGAGCTGG GCTCCTCGAG GGGTGGGGCT 420
40 GCTGTCCCTG GAAGCCTGGC AGCAGCACTG TATCGGGTTG GCTGAAGCTG ARCGCGTGG 480
GGTGCAGGGC TCCMGAATC CCCGTTGGC TGAAGGGGTT CCTGTAGCC MGGGATGTTT 540
45 ATGAGGTCTC TCTGATGCCC CAGGCGCAGG ACATGTGTGC GGGTGGAGAA AAGCAGGCC 600
TTTCAGTGCC AGCTCCACTC AATTTCTATG TGGACCAAGA ACGATAAACT TAAAAAATTT 660
TTTTTCCTAA GGTATCTTCA GAATATGGTG TATTTTTATG TGGAAAAGAA AAGTTATGAA 720
50 GGCAGCTGTT ACTTTAAGAG AAAATTCATT AAAAGTCTC GAGGTATGAA GATGACGCG 780
TGCTTCTCAA TCATTTTGGC ATAACCTGAT TGTGGCTGTA ATTTTTTTTT TTTTTTTTGT 840
55 CAAGCATGTC AGACAATAAA GTCTTTGTAA AAAGRGAAA AAAAAAAAAA AAAAAAAAAA 900
ACTCGA 906

60

(2) INFORMATION FOR SEQ ID NO: 191:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 1941 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

CTTCAGCTGA AGCCCAGGGA CCCCTTTTCC ACCCTGGGCC CCAATGCCGT CCTTTCCCG 60
CAGAGACTGG TCTTGAAAC CCTCAGCAA CTCAGCATCC AGGACAACAA TGTGGACCTG 120
15 ATTCTGGCCA CACCCCCCTT CAGCCGCTG GAGAAGTTGT ATAGCACTAT GGTGCGCTTC 180
CTCAGTGACC GAAAGAACCC GGTGTGCCG AGATGGCTGT GGTACTGCTG GCCAACCTGG 240
20 CTCAGGGGGA CAGCCTGGCA GCTCGTGCCA TTGCAGTGCA GAAGGGCAGT ATCGGCAACC 300
TCCTGGGCTT CCTAGAGGAC AGCCTTGCCG CCACACAGTT CCAGCAGAGC CAGGCCAGCC 360
TCCTCCACAT GCAGAACCCA CCCTTTGAGC CAAYTAGTGT GGACATGATG CGGCGGGCTG 420
25 CCGCGGGCT GCTTGCTTG GCCAAGGTGG ACGAGAACCA CTCAGAGTTT ACTCTGTACG 480
AATCAGGGCT GTTGACATC TCGGTATCAC CGTTGATGAA CTCAKTGGTT TCACAAGTCA 540
30 TTTGTGATGT ACTGTTTTTG NATTGGCCAG TCATGACAGC CGTGGGACAC CTCCCCCCCC 600
CGTGTGTGTG TCGTGTGTG GAGAACTTAG AACTGACTG TTGCCCTTTA TTTATGCAA 660
ACCACTCAG AATCCAGTTT ACCCTGTGCT GTCCAGCTTC TCCCTTGGGA AAAAGTCTCT 720
35 CCTGTTTCTC TCTCTCTCT CCACCTCCCC TCCTCCATC ACCTCAGCC TTTCTGTTC 780
TTGTCTCAC CTTACTCCCC TCAGGACCCT ACCCCACCCT CTTTGAAAAG ACAAGCTCT 840
40 GCCTACATAG AAGACTTTTT TTATTTAAC CAAAGTTACT GTTGTTTACA GTGAGTTTGG 900
GGAAAAAAA TAAATAAAA ATGGCTTTCC CAGTCCCTGC ATCAACGGGA TGCCACATTT 960
CATAACTGTT TTTAATGGTA AAAAAAAAA AAAAAAATAC AAAAAAAAT TCTGAAGGAC 1020
45 AAAAAAGTG ACTGCTGAAC TGTGTGTGGT TTATTGTTGT ACATTCACAA TCTTGCAGGA 1080
GCCAAGAAGT TCGCAGTTGT GAACAGACCC GTTTCACCTG AGAGGCCTGT GCAGTAGAGT 1140
50 GTAGACCCTT TCATGTACTG TACTGTACAC CTGATACTGT AACATACTG TAATAATAAT 1200
GTCTCACATG GAAACAGAAA ACGCTGGGTC AGCAGCAAGC TGTAGTTTTT AAAAATGTTT 1260
TTAGTTAAAC GTTGAGGAGA AAAAAAAAA AGGCTTTTCC CCCAAAGTAT CATGTGTGAA 1320
55 CCTACAACAC CTGACCTCT TTCTCTCTC CTTGATTGTA TGAATAACCC TGAGATCACC 1380
TCTTAGAACT GGTTTTAACC TTTAGCTGCA GCGNCTACGT CNAWCGNIGT GTATATATAT 1440
60 GACGTGTAC ATTGCACATA CCCTGGATC CCCACAGTTK GGTCTCTCTC CCAGCTACCC 1500

CTTTATAGTA TGACGAGTTA ACAAGTTGGT GACCTGCACA AAGCGAGACA CAGCTATTTA 1560
ATCTCTTGCC CAGATATCGC CCTCTTGGT GCGATGCTGT ACAGGTCTCT GTAAAAAGTC 1620
5 CTTGCTGTCT CAGCAGCCAA TCAACTTATA GTTTATTTTT TTCTGGGTTT TTGTTTGTGTT 1680
TTGTTTCTT TCTAATCGAG GTGTGAAAAA GTTCTAGGTT CAGTTGAAGT TCTGATGAAG 1740
10 AAACACAATT GAGATTTTTT CAGTGATAAA ATCTGCATAT TTGTATTICA ACAATGTAGC 1800
TAAAACTTGA TGTAATTCC TCCTTTTTTT CCTTTTTTGG CTTAATGAAT ATCATTTATT 1860
CAGTATGAAA TCITTATACT ATATGTTCCA CGTGTTAAGA ATAAATGTAC ATTAATCTTT 1920
15 GGTAAGACTT TAAAAAAA A 1941

20

(2) INFORMATION FOR SEQ ID NO: 192:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2118 base pairs
25 (B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

30 AAATAATAAT AANAATAAAT AAAAATWAAG TGCTTAKTGT AACTCAGCGG ACAGGGCTCC 60
CAGCTGCTCT GGCACGTGGG ACACCTTCCA CCTGACAC AACAGGCATG CAAAGAGGAC 120
35 TGGATATGGT GGGGTAGAGT GCTTCTGGTG TGTTCACCTT AAGAAAACAT CTGCCAAGAG 180
AGAAGAGTGC CCAGGAAAGA CCAGGAAAT ACAAGTACAT GGCTGCTTCA TACCATATAC 240
CCCAATTCCT TAAAGCAGCA AAAGGCACTT TTTTTTTCAG GCCAGAGTGA ATCTAAAACA 300
40 AACCTGGCTT TGCTTACAGG GAAGCTGTCC CAGAAGGACT GAGTGATGCC TCTTGTCCC 360
TAAGTCTGG AGAGTCTTTG CAAGTTTCCA ACGACATTTC CAACCAGGTG GGAGAGACCA 420
45 GCAGTTGACG AGACAAGTCA GACCCAAAAA ACGACGCCAA GGTAGTGAGT GGGTGCCTAT 480
TTGGGAGTAG GATGATTGA GGAAACAGG AAGAAAAACC GGTGAGAAAG TGGCACTTTG 540
GAAGTGAAA GCTGTTTGA AATAGCAACT CTGGCTAAAG CGAAAATGTT AATCAAGTAG 600
50 AAAGTAAAT TCAGGATCTT AGAAGCTCAT CCTTCTGATG AGAACTATTT TTTTTCCGT 660
GAAGGAACTA TTATTACTTT AAAAGTGAGG GTAATTTACA TATGGGGTGT ATATATTCTA 720
55 AAAATAGTAA TAAAGTACC TTTTATAAGC AATGTTGTGT GGCTTGTAAG AGAAAGCAGG 780
GAGGAAAAA AGGCAGGCAA AACTAGTCTA GGTCTAGGCC CTAAAAATGA GCTTCCTTCC 840
CACTTGACTG GAAACGCCCA TGTGATTTCT AGGCTGAAAA TAGGTAGGAT TTAACGAGTA 900
60

ACCTAGTTCC CTTCTGTCTC TGATTCTGA TCAGCTGATG GAGCTGCTAG TAAGAGGGGC 960
 CGATCATGCT CCCAGACGAG TCCTTTGGCC TCTTGCTCTC CATCCCAAGC CTGACTCCTT 1020
 5 CAGCAGCAGC CCCCTCCTTC TGTGTCCATC TGATGCAGGC AAGCAGGAGC AGTAAGAGGG 1080
 CATCCCATGT TCCAGTTCAC CTTCTATGGG GTGACTARGA GGTTCCTGGT AACTAGGGCA 1140
 10 GCCCARGCCC AGCAGGTTCG AAAAGCAGCT GCAAGCTTCA GAAACCCACT TCCTCCAACA 1200
 CCAGGGAGGT GGCAGAGAGC CCATCCAAAA GCCCACTGGG AGAGGCATAA GATTCTGTGC 1260
 CAGGCCCCCA GGTCCCCTCT GTGTCAGGTA GGCTCTGCTA CTGGCCTCTG AAGTAAAGGC 1320
 15 AAANACAAAC GGGCAGGGCA GGTGGCAGG AATAAAAAAC TCTGGACAGA AACCTTTTA 1380
 ATAAAGGAAA TTCCACCCCT CCCAATCCTT CCATGGAAGG GTGAGACCTT AATGTGATGT 1440
 AAGAGGAAGG TCTTCTCTGG CTTTCAGGGA AACAGCTGCA GCTGAAACTT AGGGGCCCAT 1500
 20 TCCAGGGCAC TTTTCACCAC AGCCAGTGCA GCCGCTCCAA GTGCCACTGT CAGCCCCATC 1560
 ACTGCCAATT TCACAAAGCG GTTGGTCTTT GGCTTGGTCA GGACATCTTT TGTTCGATCT 1620
 25 TCAGGCCGCA GAAGTCCCCG AANACCGCTG CGCAGCACC ATATCAGGCC TCTGCTGGGC 1680
 TGATGCCAGC TCAAAGTCTT TGAAAGTAGA GGCTGCCGTC CTCTCAGCTT GCTGTTGGGC 1740
 AGCGGCCTCC CGAGCAAGTT CGGATGGGGG AAACTGAACA AAAAGGTCTC CTSTCTGCTG 1800
 30 ATCAGTGTCT CATAGGGCAA GTCCTGAGGG ATCTGGGACA ACAGGTGGTG GACCGAGGCC 1860
 ATGTCACAGT CACAGTCCAG GACTTCCTGC TCGGATACA ACACAATCAC GGCTGCAAAG 1920
 35 TAAATCGGCA TCACTGGGTG GCAGGCCAGG AAGAAGTCAT ATAACCGCAC GACGTGCCTG 1980
 AAGTCAGACA GGACATGCCC AAACCAGGTG ATGAGCCAGC TGAGGGCAAA GATGGTCCCT 2040
 ACCTCAGCAC TCTGCATGAA GTCATGGAGC TCTGGATTCA CCTGGTCAAT GATGGGCATC 2100
 40 AGATAGTTTA ATATATGC 2118

45

(2) INFORMATION FOR SEQ ID NO: 193:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1538 base pairs
 50 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

55 CCGGGTTCGG CTCTGTGTCA GCAGCCGGGC GGCGCTGGG CCGGACATGG CAGCCTGTAC 60
 AGCCCGGCGG CCTGGCCGTG GGCAGCCGCT GGTGGTCCCG GTCGCTGACT GNGGCCCGGT 120
 60 GGCCAAGGCC GCTCTGTGCG CGGCCGNAGC TGGAGCCTTC TCGCCAGCGT CGACCAGGAC 180

	GACGCGGAGG CACCTCTCGT CCCGAAACCG ACCAGAGGGC AAAGTGTGG AGACAGTTGG	240
	TGTGTTTGAG GTGCCAAAAC AGAATGGAAA ATATGAGACC GGGCAGCTTT TCCTTCATAG	300
5	CATTTTGGC TACCGAGGTG TCGTCTGT TCCCTGGCAG GCCAGACTGT RTGACCGGA	360
	TGTGGCTTCT GCAGCTCCAG AAAAAGCAGA GAACCTGCT GGCATGGCT CCAAGGAGGT	420
10	GAAAGGCAA ACTCACACTT ACTATCAGGT GCTGATTGAT GCTCGTGA CTGACATAT	480
	ATCTCAGAGA TCTCAGACAG AAGCTGTGAC CTTCTGGCT AACCATGATG ACAGTCGGGC	540
	CCTCTATGCC ATCCCAGGCT TGGACTATGT CAGCCATGAA GACATCCTCC CCTACACCTC	600
15	CACTGATCAG GTTCCCATCC AACATGAACT CTTTGAAAGA TTTCTCTGT ATGACCAGAC	660
	AAAAGCACCT CCTTTGTGG CTCGGGAGAC GCTAAGGGCC TGGCAAGAGA AGAATCACCC	720
20	CTGGCTGGAG CTCTCCGATG TTCATCGGA AACAACTGAG AACATACGTG TCACTGTCAT	780
	CCCCTTCTAC ATGGGCATGA GGGGAAGCCCA GAATTCACAC GTGTACTGGT GGCCTACTG	840
	TATCCGTTTG GAGAACCTTG ACAGTGATGT GGTACAGCTC CGGAGCGGC ACTGGAGGAT	900
25	ATTCAGTCTC TCTGGCACCT TGGAGACAGT GCGAGGCCGA GGGTAGTGG GCAGGGAACC	960
	AGTGTATCC AAGGAGCAGC CTGCGTTCCA GTATAGCAGC CACGTCTGC TGCAGGCTTC	1020
30	CAGTGGGCAC ATGTGGGGCA CGTTCGCTT TGAAAGACCT GATGGCTCCC ACTTTGATGT	1080
	TGGGATTCCT CCCTTCTCCC TGGAAAGCAA TAAAGATGAG AAGACACCAC CTCAGGCCT	1140
	TCACTGGTAG GCCAGCTGAG GCCCCAAGTG CCCAGGCTTG GTCACCGGGA AGAACAATC	1200
35	TCATCCACA ATTGCTGCAG AACTCTTCTC TCCCATCAT GGGCCACAGT GGTCTCTTA	1260
	ATTTGATTGT GGGTTCTTT TTGTGGGAG GGGTGTATA ACTTTCTTC AGAAGACCCA	1320
40	TGTGGGACAC CTCCAAGGCT GGCCTCTCA TAAGCCCTGC CTACACCATG TTCCAGTAAA	1380
	CCTCTCCACC AAGGAACTGT GTTCAGCTGC CACAGGCTG GAGGAGTTTC CTGGCCTGTC	1440
	ACGTGAGGTT TGATCAGTAA ACCAGTGCAS GYTTGGCCAA AAAAAAAAAA AAAAAAAAAA	1500
45	AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAACCTGA	1538

50

(2) INFORMATION FOR SEQ ID NO: 194:

(i) SEQUENCE CHARACTERISTICS:

55

- (A) LENGTH: 1098 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

60

AGACCCCTGTC TCAAATAATA ATAATAATAA TAATCTTATT TTGGAGAATA AAGAGACCTS 60
 TGGATTGTAG GTGCCATTTG GGTAGAAAGA AAAGACGTTT ACACCGAGAA ATAGTCTGTG 120
 5 TTGCCCTGAA GGAGCAGAGG GATGCATCGC TGGAGGTGAC CTACAGTTGA AGAAGACTCA 180
 TTATGACAGA CCTGTCTCTT CTTCCTTGTG GAAAGTGTTT CCTCTGCTGC TACTGCTCAT 240
 GAGACTCTTC CCCCTCCTTG TCCCAGGGAA CCAAAGGCT TTNCTACCAC ACCCTTTCTT 300
 10 NGCCCCCGC CTCCCATGTC TGCTGTGCCT TTGTACTCAG CAATCTTNG TTGCTCCCA 360
 TTATCTTCCA GCCGATACA GAGTGAATAG TTAACCACAC TTAGGTCAA TAGGATCTAA 420
 15 ATTTTGTTC CTGCTCCNGT GTAAAGAGGC CAGTGTGTTG GTGTTGCAAG CAGCCTTGA 480
 ATAGTAACTC TTCTCATTTG TTTGGGATCT GGCCAMCAAG TTCCAGAATG ATACACGGAT 540
 CAGTGCAGAA GTTCATCAGG CTCTCGGACC TTAGGGCTGT TGGAGAAGGC TTCAGCAGCA 600
 20 GAACTGATGG TKAWKGYTCG TGTCTCCAT CCTCAACTTT CTMTGCTTCG ATCATAACACA 660
 AGAATACATT TGGAAGGGCA AAAAATGAAC ACTGTGTTC ATTCAGCCG TGTTTGTGA 720
 25 CACAGATGCA CAGTCTGCTG TGAAGACCTT CTCTCAAGTG GSATYTGGA GTCCATGCCA 780
 GATCATGGTG CTTCATGAGA GACTGACAGC TATCAGGGT TGTGGCACTT AGTGAGGACT 840
 CTCTCCCCC AGTGTGTGCT GATGACACAT ACACACCTGA CAATAGCTTG AGTCTTCTCT 900
 30 GTTCCTTTTA CTCTGTAGCC AACATACACA TGATTTAAAA CCCTTTCTAA ATATCTATCA 960
 TGGTTCATCC TTGTCCAAAT GCAGAGTCAG AGCTATTGTT ACTTCATTAT TATTTCCAAG 1020
 35 GCGAATAGTT GGCTTTCTTT TTGCAAAAAT AATTAAAGTT TTTGTATGTT GCAAAAAAAA 1080
 AAAAAAAA CTACGTAG 1098

40

(2) INFORMATION FOR SEQ ID NO: 195:

45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1001 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

GAATTCGGCA CGAGATAGCT TGCATCTCAT CCCAGTAAAA CCACCTATTT ATAACATATC 60
 AACGTATTGA CAAGTTGAA GAGCAAGATT GTTCTGAGGT GAGATGCAA TTTCAAAGG 120
 55 GTGAGCACTA ATTGTCCAG TGATTGTTTA TTTATTGGCT AGGACATAAT TACTCTCTTT 180
 GAGGTTACAC ATCTGCCTCC AGGTTCTGT GTGCTGTGC CCTTGGGATC AGGCCAGGGC 240
 60 AGACTGTGAT CACTGAGATT CAACTCCCA GARTAAATCAG CAAGAGCTTT CTAGAGACCA 300

5 AGGCCAGGCC TGATCCCTGA GGGATGCATG AGAAGGCTTG GAATCTCATT CTGCTATGGT 360
 GGCTCTCTCT TGATCTTCTT GGAGTAGCAA AAACAGCAAT GTGGGCCCAA TGGTGTGGCC 420
 TAAATGATCA CAAAGGTAAA TGAGTAAAGG GCTCAGCAGA TGAGTAAGGA GCCTTGTCTT 480
 GAGAAATTAG CACTGGGCTC TGCATTGAGA AACATGTGAT AAGCATTGCC CATTGCACAT 540
 10 TGCCTTTATT GTGTAAGGAC ATGAAATTCC AGTTTTGCAT AGCTAGTGAT GAATACCTGA 600
 AGGGAATTGC AGACATATTT TATTTTATTT TTAATTGACA GATGGAATTG TATATATTTA 660
 TCATGTACAT AATCATGCTT TAAAATATGT ACATTATGGA ATGGCTAAAT CAAACTAACC 720
 15 TAGGCATTAT CTCATATAAT TGTCATTTTT GTGGCGAGAA GACTAAAAAT CTACCCCTTC 780
 AGCATTTTTA AAGAATACAA TGTGTTTTAT TAACAACAGT CACCATTTGG TACACTAGAT 840
 20 CTCTTGAAC TCTTCTCTT ATCTAACTGA GATCTTGTA CCTTTGATA CAGCTCCCAA 900
 GCCCTTCCCC AACCCTGCT CCACCCGTGG TAACCACCAT TCTATTCTCA ACTTCTGGT 960
 25 AATCACCATT CTAGACACAG GGAAGACTCT CTACCCCTCG A 1001

(2) INFORMATION FOR SEQ ID NO: 196:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1443 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 196:

40 ATAAACTGAA ATAGGTCATG CAAATATAAA ATATTATTTT TAAATTATTT GTCATAAGAA 60
 ACGATGGTGG CCATATTTTG CTTTAATAAT GGAAAAATG TGGTTAGCAT TCTKTGGAAG 120
 GTGGTCATCA GATAGTAGAC ATTTTCTAGG ATTATTTCT ACCTGCATAT GTGGAAATGT 180
 45 GTACTACTTT AGATTATWT AATGGCAGCT AACTCAGAGG CATCAAAATG TGCTAATGGT 240
 GTAATATGGC CTTTGTCTTG CTGTYCTGTT TTGTARGCCT TCAATCAAGC ARGGGCAGGG 300
 CCGTACAGTG AACTTGTCTT TTGSCAGACG CCAGCGTCTG CCCCTGACCC CGTCTCCACT 360
 50 CTCTGTGTCC TGGAGGAGGA GCCCCTTGAT GCTACCCCTG ATTCACCTTC TGGTGCCTT 420
 GTACTGAACT GGAAGAGCC GTGCAATAAC GGATCTGAAA TCCTTGCTTA CACCATTGAT 480
 55 CTAGGAGACA CTAGCATTAC CGTGGGCAAC ACCACCATGC ATGTTATGAA AGATCTCCTT 540
 CCAGAAACCA CCTACCGGTG AGTGCAAGGG AGTAGAAATC TGCATCAGCA CATCAGCACT 600
 60 TGGGGATCTA AGTAAACCTC TCGGGGAAA TGACCAAGTG GATGTCATCT CCCAGCTGTT 660

TCTAAGAGCC CAGATGTCCA GAGTATGTGC TCACCTTGAT COCTCAGGCC AGAAGACCTG 720
 TGAAAAAGCC ACACCTGGTC AGGGACTCAC TGGACGGTTT TGTGTCCACT TTAACCTGCA 780
 5 CCGTCTCTAC CCCAGAGTGG ACTCARATCC TCACCTGATC CTCTGACAT TGGCTGCA 840
 AATTATAAAA GGGCTTTGGC AATATGTTAG CCCAAGAAAT TGGCTTCTTC CAGAATTTG 900
 GCCGACNMTA ACAGTGGCTT AAATGATGGT AAACCTTTTA AGATTCTTAA AAGCTTGGCA 960
 10 TTGGAGATAC GTTGACTTTT ATTAAACMAC CTATGTTGT TTAATGATTT CTAAAAAAT 1020
 ATCTGGAGCT CAGGGGTTC ACTGAGGGAA CACATTTGA GRATCATTT TTAATAATTA 1080
 15 AATGCCAGGT AACCCGTGA AATTATCAAA AACACCTTC ACGTACCAGA AAGTACCTCA 1140
 GAGGATAGTT CTGTTATGGA GAAGATGAAA TGGTTTASTA GTGTAGGAC TATGGAAAGG 1200
 TGAGCTTAGA TTTGATAGT AAAACCTCAA GACCTTCTTT AAAAGTATT TTAATGATGC 1260
 20 AGCATAAATA ATTAAATCA GTGTTAANAT GCCAAGCTTA GTATATTGAG CTGAATGTGA 1320
 AAAGAACTC ACATTGGAG AATGCCACCT TTCTCTATA AGATAGCTTT GAAGTACCA 1380
 25 TTTTAGACAG ATGGAAATG AATAGCTTTA GAAAGGCCAA ATGTTTGATC TTGGGGAAAA 1440
 AAA 1443

30

(2) INFORMATION FOR SEQ ID NO: 197:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1282 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

GAAAAAAGAA AGTATGACCC AGTAGCTAGG CACCTTGGC CCGCCAGT TGCACATAA 60
 AATTAACTGT CACAGTATCA TCTTAGAAGT GAAAGAGCC CTTTATCTT GGTGGCCCC 120
 45 TCTACCACCA CTTACTGACA AAGAACATGG TGCTATCTGG CATGGGAGAA ATTTCAGTT 180
 TGCTATGGCT TGTATGTGC COCTCAAAT CAAGTGTGC CAATGTGCA GCATCAAGAG 240
 50 GTGGGTCTT TAAGAGATCA CTAGGCCATG AGGATCTC TTAGGACTGG GATGAAGGCC 300
 CATAATAAAA GAGGTTTCAG GGAGCATCTT GCTAGCTGC CTTCTGTATG TGAGAACACA 360
 GCAAGAAAGC CTTAGTCAAC AAGTCCAGC TCTTGTCTT TAGACTTCC AATCTCCAGA 420
 55 ACTGTGAGAA ATACATTCT GTTCTTACA AATTCCAG TCTCTGTAT TTTTATATAG 480
 CAGCACAAA TGAAGATACC ATACCTGAC ACCTGAACAT TCTCACAG GTAGTAAATG 540
 60 CACTGCTTTA TTCTGTCTC AGTATTGTGT GCTTAAAG GAAATGAGAA AGGTGGATC 600

450

AGGGCATAGG ATGAACAAGT TACTGCTAGA CCTCTCAGAA TGCCACTAAT GGAATAGATT 660
 GTATTTTCAT CATNCTTGT CTCTTCGGAA GCTAACACCA TGTATATAA GGCATTAAT 720
 5 AGATGTCTAA AACACCTTA AGTATTGTG TAGAAATCTG GTGCATGTC CAGAAAGAAC 780
 CAAAATTCMA AATAATTICA AAGGGCCTAA AGCACTATTT AATCMATTT CATTAGTTTT 840
 10 TAATGGTACT ACCACTCTCA AATTTAAAAT GTGATCTTAC GTTCCTCTTC CTCGATTTGG 900
 ATTTATTGCT AAAACCTGGT AAACACTTTA ATCCYTTTCA ATCCCATTC CACTGCTCTT 960
 GTCCAGAATT ACTCGCAGAC TAATAGTCAC CTGACTTCTC CCGCTGCTTC CCGATTGTCT 1020
 15 GTCTAATCTT GGTACAAAT AAGTAACTGC CAACTAATC TTCTAATAA GCAAGACTGA 1080
 TCTGCTCACT CTTTGCTCA ACAATGTAAA AGCTCCCATT GTCTCCCAA TAAACCCAGC 1140
 20 TTTCCACTGT GTATACAATA CATCCATGAT CTGTATCCAG CATCATTTTG TATTGCTCA 1200
 CTTTATACAC CACCCCCCAT GCCACATCAA ATTAAATTAT COTGATTAAT GCAATGCAA 1260
 AAAAAAAAAA AAAAAAATC GA 1282
 25

(2) INFORMATION FOR SEQ ID NO: 198:

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 951 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

35

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

ATTTCCGAAC GAGGACTGAA GTGGGAGCGG CGGCAGGTA GAGACAGAA GGGGATCTA 60
 40 TGTGGTAACT AAAGAATGTT TCTGTTTGT TAATTATGTT GTGTGTGTTG TTTTATGTT 120
 TGCTTAAGAG AATCAAAAAC TGAAAAAAT GAGAATACAG GAATGGCTC TGGTTATTT 180
 45 TTTTGCTGTG TTTACAGCTT GTTAATGCTC TACTGTCTTT GTTCAAGAG AGATTGTTC 240
 ACTGCCCAGC TGTMTTGTG TCTGAGCCC TATGCCCAGC CCACCTATA AATCTGCCT 300
 GTTTAGATGT TTGATTTTGT TCTGTTTGT ATTGTTATCT TAAAGGTGA TACTCTGAC 360
 50 ATGCCAGACA TCAAAATTAAG CTCAAAATTA GCTCTCGTTT AATGTTTAA ACCTTAATT 420
 TATAATCTAA TTGATCCAG CCACTGATGC ATGTACTTTA GCTACTTTG CTAAATAAGC 480
 55 ATATTAATTT TCCACATCAG GCCATCAGAT CTTGAGAACC AACAGTTATC TAGAATCCG 540
 TGTCTACTAA TGTTTCACCT GCATGCAGCC TTCATTAAT TGTAGCAA ATATAAAGTG 600
 ATCATTATGT AGTTTCTGGA TAAAAAAT TTGTGTGTA AGTTGCTTTG TAAGTGCAAT 660
 60

GTGGAATTAA TGGGACAGTG TGCCCTTTGT GTTAGATGTT AGAGCAAAAG AAAGGGCTTA 720
 TAGTGTTAGT ATTGGAGCAC TTTGAAGATA GATATTTTCA GAAAAGATGT AGGATTTAAA 780
 5 AGTTAAATTT TAAATTTTAG AAAAAGATAT GATGGCAATT GGAAATAGTC ACAATGAAGT 840
 TCTTCATCCA GTAGGTGTTT AACAGTGTTA TTTTGCCACT GGTAATGTGT AAACGTGAG 900
 TGATTTACAA TAAATGATTA TGAATTCAAA AAAAAAAAAA AAAAAACTCG A 951

10

(2) INFORMATION FOR SEQ ID NO: 199:

15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1740 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

TTATTATAAT AATGATGATG ATTCCAAGGA AAAAACCTAC AGCGAATGTT CCATTCTCTAC 60
 25 CCGGCACGCA GACACTCTCC CTAACACTGA TAACCTGAGC CCCCAGCACT GGACGGAAGA 120
 ATGCTGGCGT CTCGGTGTGT ACTGGTTCAG GGTCTGGCC CCAGCCTTGT CAGGACCCCC 180
 30 TGGTGTCAG AGCCCCCACC CCTCCGCAA CAAGCAGCTG ATGCCCCAGT GATTCTCTAT 240
 ACATTTTCA CCTGGCCAA TATGTCCAGG AAAACTGCTT ACTTCTCTTT TCTTGCTGG 300
 AGCCTTCATT GTTCACCTT ACGTTGCAAT ATAGGAATTA ATGCTACAAA ATAAAAGTAA 360
 35 AGCTTACCTG AAAAGTGCAT AGTTTGGGGC AATGGTATCT ACATCTCCCA CTGTGGGAAA 420
 ACCAGCAAAG CATCAAACT CTCAATTCTC CTGTTACCRA ATGCAGATCT GAATTATAAG 480
 40 ATGTTTATGT TTGACCATG TTTCAACAAT GGGATTTTGT TACGAATTAT CCTTTAACT 540
 GAAACCTCA GTTTTACTGT TTACATTATT AGGAAAACAG GGATATCTTT TGAATCTAAA 600
 AATTGTATGT ACAGCATGTG ATTTTGAAG TTTACATGTA AAGTCACAGT ATAGGTGAAA 660
 45 TAACGTTTGT CATATTTTGA GACGTATCCT GCAGCCATGT TTTTACGTGA GTGTTTTAGT 720
 CAAAGTACAT GGTAGACAGT CTTTCACAAAT AAAAGGAAAA GGATTTTTTT TCCTCCAAAT 780
 50 GTACATTTAT CAACCTAATG ATTGATTTTT TAAAAAGAG ATTTGCCCCC AGTCTGGTTT 840
 ATGAAAGTTC ATTGCCCTAA ACTGTGCTGA TTGTTTTTAA TCAAGTTATA AATTCCAAC 900
 CTAGATCATG TATCTACCAA CTCTCTGCA TTTTCCAAA GGCATTGAGC TTAATATTA 960
 55 GTCTTGCTTA GAGTAGGTTA TCCACTTACA TGCTGCGCTA AAGCCATGCC TTTGAAACTC 1020
 CTTGTTTAAA ACATGATATG ATTTTGTGG GCAGTTTCAG AAAAGAAAAC AAACAAACAA 1080
 60 AAATCGACCC TTTAATTATT ACTTGCAACT CAACAGATCT CCCTGCCGTA CTGCCTTTTC 1140

5 CAGGAACITTT ACTTCAGGGC TGTCAGATT GCAGTTGTGC CCCGTGTATG TGGATCTAGT 1200
TCACAGAGTC TTTGGAAGCC AGCAGTCGTG CCCTCCGTAT ACTGTCCACT CATTTTATGT 1260
AGATTTGGTA TCCTCAGCAG CCAGTGTTAA CACCACTGTC ACGTAGTTAN CAGATTCATC 1320
TTTTATGTAT TTAAAGTAAT CCATACTATG ATTTGGTTTT TCCCTGCACC ATTAATTCTG 1380
10 GCATCAGATC AGTTTTTGTG TTGTGAAGTT CTA CTGTGGT TTGACCCAAG ACCACAACCA 1440
TGAGACCCCTG AAGTAAAGAT AAGGTACACA TACATTATTT GAGTAACTGT TTCCTTGGGG 1500
GCCAATCTGT GTATGCTTTT AGAAGTTTAC AGAATGCTTT TATTTTTGTC TATAACAAAC 1560
15 AGTCTGT CAT TTATTTCTGT TGATAAACCA TTTGGACAGA GTGAGGACGT TTGCCCTGTT 1620
ATCTCCTAGT GCTAACAAATA CACTCCAGTC ATGAGCCGGG CTTTACAAAT AAAGCACTTT 1680
20 TGATGACTCA MAAAAAAAAA AAAAAAAMC YCGGGGGGGG GCCGTAACC CATTTNNCCC 1740

25 (2) INFORMATION FOR SEQ ID NO: 200:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1707 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

35 GCTTATAGAA GGGAGAGGAG CGAACATGCC AGCGGTTGG CGGTTTGGT GTGTCTCTGT 60
GACCATGGTG GTGGCGCTGC TCATCGTTTG CGACGTTCCC TCAGCCTCTG CCCAAAGAAA 120
GAAGGAGATG GTGTTATCTG AAAAGGTTAG TCAGCTGATG GAATGGACTA ACAAAGACC 180
40 TGTAATAAGA ATGAATGGAG ACAAGTTCCG TCGCCTTGTG AAAGCCCCAC CGAGAAATTA 240
CTCGTTATC GTCATGTTCA CTGCTCTCCA ACTGCATAGA CAGTGTGTCG TTTGCAAGCA 300
45 AGCTGATGAA GAATCCAGA TCCTGGCAAA CTCTGGCGA TACTCCAGTG CATTCACCAA 360
CAGGATATTT TTGCCATGG TGGATTTTGA TGAAGGCTCT GATGTATTC AGATGCTAAA 420
CATGAATCA GCTCCAACCT TCATCAACTT TCCTGCAAAA GGGAAACCA AACGGGGTGA 480
50 TACATATGAG TTACAGGTGC GGGGTTTTTC AGCTGAGCAG ATTGCCCGGT GGATCGCCGA 540
CAGAACTGAT GTCAATATTA GAGTGATTAG ACCCCCCAAT TATGCTGGTC CCCTTATGTT 600
55 GGGATGCTT TTGGCTGTTA TTGGTGGACT TGTGTATCTT CGAAGAGTAA TATGGAATTT 660
CTCTTAATA AACTGGATG GGCTTTTGCA GCTTTGTGTT TTGTGCTTGC TATGACATCT 720
GGTCAAATGT GGAACCATAT AAGAGGACCA CCATATGCCC ATAAGAATCC CCACACGGGA 780
60

453

CATGTGAATT ATATCCATGG AAGCAGTCAA GCCCAGTTTG TAGCTGAAAC ACACATTGTT 840
 CTTCTGTTTA ATGGTGGAGT TACCTTAGGA ATGGTGCTTT TATGTGAAGC TGCTACCTCT 900
 5 GACATGGATA TTGAAAGCG AAAGATAATG TGTGTGGCTG GTATTGGACT TGTGTATTATTA 960
 TTCTTCAGTT GGATGCTCTC TATTTTATAG TCTAAATATC ATGGCTACCC ATACAGCTTT 1020
 CTGATGAGTT AAAAAGGTCC CAGAGATATA TAGACACTGG AGTACTGGAA ATTGAAAAAC 1080
 10 GAAAATCGTG TGTGTTTGAA AAGAAGAATG CAACTTGTAT ATTTGTATT ACCTCTTTT 1140
 TTCAAGTGAT TTAAATAGTT AATCATTTAA CCAAGAAGA TGTGTAGTGC CTTAACAAGC 1200
 15 AATCCTCTGT CAAAATCTGA GGTATTGAA AATAATTATC CTCTAACCT TCTCTTCCCA 1260
 GTGAACTTA TGGACATTT AATTTAGTAC AATTAAGTAT ATTATAAAAA TTGTAAACT 1320
 ACTACTTTGT TTTAGTTAGA ACAAAGCTCA AAACACTTTT AGTTAACTTG GTCATCTGAT 1380
 20 TTTATATTGC CTTATCCAAA GATGGGGAAA GTAAGTCTTG ACCAGGTGTT CCCACATATG 1440
 CCTGTTACAG ATAACTACAT TAGGAATTCA TTCTTAGCTT CTTTCATCTTT GTGTGGATGT 1500
 25 GTATACTTTA CGCATCTTC CTTTGAGTA GAGAAATTAT GTGTGTCATG TGGTCTTCTG 1560
 AAAATGGAAC ACCATCTTC AGAGCACAG TCTAGCCCTC AGCAAGACAG TTGTTTCTCC 1620
 TCTCTCTG ATATTTCCTA CTGAAATACA GTGCTGTCTA TGATGTGTTT TGTGTTGTTG 1680
 30 TTTTTFYAG ATCAGYTAC TGGGCTC 1707

35

(2) INFORMATION FOR SEQ ID NO: 201:

(i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 779 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

45 CTGTCCCCAG TGTTCOCAG TAATGACTTG GCACTCCAGA GAAAGTTTCA TRCTGTTGCG 60
 TGTGGTGGCT CCAAGCCAAG CACCTGGCAT GCAGTCCAGC CCTTCCAGC GGGCGTGGCG 120
 50 TCGTCTCTT CACAGATGCC ACGTTGCAGC CCAAGGCCT CACCATTTTG CGTTTATTAG 180
 AAACCATTT TCTTGGTCAT TTATAAGCT GCTTTATAGA TATCTTTGAT CCTGGCATGC 240
 CTGGTTTCC TCTCCCTCC CTCTTTCCAA TCTGGTTTC CTAACCTCCT CTGTAGTAA 300
 55 TTCTCAACTC AACTCAAAGT CCAAGAATT TGAATGGTA GGATGCTGTG CGGGAGCTC 360
 GAGGCTGAGG CATAATCACT GCTTCGGTTC TGCTCATCAG GGGACACGCT CCCTTACTCA 420
 60 TGGCAGCCAT GTTTGATTGT CACAGAGCCC CCCGAATACT CTGTCTATAG TGACACACTG 480

454

TAGGTGTCAT AAATTTTAAG AAACCTGCTT TTAAGTACTA TTTATAGGTT TTTCTGTTAT 540
 ACTTGCAACC TAGTTTTTAAA ATACATGAGG ATTTTATGAA AGCTTTATAC AGACATTTAT 600
 5 AGGAAACTCA TTCTTTGATT TTAGGTGCCA TTTAAATTGA TAACACTTAC TTTATAAAAA 660
 GATGCTTTTT GTCTGGATAG AGCCTTATAG TTTAAAATAT CTTCATATAT TGCCATTGTA 720
 10 TCAAATAAAT TTCTTACTTA GAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAACTCGA 779

15 (2) INFORMATION FOR SEQ ID NO: 202:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1617 base pairs

(B) TYPE: nucleic acid

20 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

25 GGCACAGCTT TCCTGCTCTT CCTCGCTCCC TCTCTTTCTC TCCTCCCTCT GCCTTCCCAG 60
 TGCATAAAGT CTCGTGCTCT CCCGGAAGTT GTTGGCAATG CCTATTTTTT GGCTTTCCCC 120
 CGCGTTCTCT AAATAACTA TTTAAAGGTC TCGGTCGCA AATGGTTTGA CTAAACGTAG 180
 30 GATGGGACTT AAGTTGAACG GCAGATATAT TTCACTGATC CTCGCGGTGC AAATAGCGTA 240
 TCTGGTGCAG GCCGTGAGAG CAGCGGGCAA GTGCGATGCG GTCTTCAAGG GCTTTTCGGA 300
 35 CTGTTTGCTC AAGCTGGGCG ACACATGGCC AACTACCCGC AGCCTGGGAC GACAAGACGA 360
 ACATCAAGAC CGTGTGCACA TACTGGGAGG ATTTCCACAG CTGCACGGTC ACAGCCCTTA 420
 CGGATTGCCA GGAAGGGGCG AAAGATATGT GGGATAAACT GAGAAAAGAA TCCAAAAACC 480
 40 TCAACATCCA AGGCAGCTTA TTGAACTCT GCGGCAGCGG CAACGGGGCG GCGGGGTCCC 540
 TGCTCCCGGC GTTCCCGGTG CTCTGGTGT CTCTCTCGGC AGCTTTAGCG ACCTGGCTTT 600
 45 CCTTCTGAGC GTGGGGCCAG CTCCCCCGC GCGCCACCC ACACTCACTC CATGCTCCCG 660
 GAAATCGAGA GGAAGATCCA TTAGTTCTTT GGGGACGTTG TGATTCTCTG TGATGCTGAA 720
 AACACTCATA TAGGATTGTG GGAAATCCTG ATTCTCTTTT TTATTTGTTT TGATTCTCTG 780
 50 TGTTTTATTT GCCAAATGTT ACCAATCAGT GAGCAAGCAA GCACAGCAA AATCGGACCT 840
 CAGCTTTAGT CCGTCTTCAC ACACAAATAA GAAAACGGCA AACCCACCCC ATTTTTTAAT 900
 55 TTTATTATTA TTAATTTTTT TTGTGGCAA AAGAATCTCA GGAACGGCCC TGGGCACCTA 960
 CTATATTAAAT CATGCTAGTA ACATGAAAAA TGATGGGCTC CTCCTAATAG GAAGGCGAGG 1020
 AGAGGAGAAG GCCAGGGGAA TGAATTCAAG AGAGATGTCC ACGGACGAAA CATACGGTGA 1080
 60

455

ATAATTCACG CTCACGTCGT TCTTCCACAG TATCTTGTTT TGATCATTTC CACTGCACAT 1140
TTCTCTCAA GAAAAGCGAA AGGACAGACT GTTGGCTTTG TGTITGGAGG ATAGGAGGGA 1200
5 GAGAGGGAAG GGGCTGAGGA AATCTCTGGG GTAAGAGTAA AGGCTTCCAG AAGACATGCT 1260
GCTATGTCA CTGAGGGGTT AGCTTTATCT GCTGTTGTTG ATGCATCCGT CCAAGTTCAC 1320
TGCCTTTATT TTCCCTCCTC CCTCTGTTT TAGCTGTAC ACACACAGTA ATACCTGAAT 1380
10 ATCCAACGGT ATAGATCACA AGGGGGGAT GTTAAATGTT AATCTAAAAT ATAGCTAAAA 1440
AAAGATTTTG ACATAAAGA GCCTTGATTT TAAAAAAGAG AGAGAGAGAG ATGTAATTTA 1500
15 AAAAGTTTAT TATAAATTAA ATTCAGCAA AAAAGATTG CTACAAAGTA TAGAGAAGTA 1560
TAAATAAAA GTTATTGTTT GAAAAAAGAA AAAAAAAGW CTCGACCGCA AGGGAAT 1617

20

(2) INFORMATION FOR SEQ ID NO: 203:

(i) SEQUENCE CHARACTERISTICS:
25 (A) LENGTH: 1974 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

GAATTGGCA CGAGGCTGAG GGAGCTGCAG CGCAGCAGAG TATCTGACGG CGCCAGGTTG 60
CGTAGGTGCG GCACGAGGAG TTTTCCCGGC AGCGAGGAGG TCCTGAGCAG CATGGCCCCG 120
35 AGGAGCGCCT TCCTGCGGC CGCGCTCTGG CTCTGGAGCA TCCTCTGTG CCTGCTGGCA 180
CTGCGGGCGG AGGCCGGGCC GCCGAGGAG GAGAGCCTGT ACCTATGGAT CGATGCTCAC 240
40 CAGGCAAGAG TACTCATAGG ATTTGAAGAA GATATCCTGA TTGTTTCAGA GGGGAAATG 300
GCACCTTTTA CACATGATTT CAGAAAAGCG CAACAGAGAA TGCCAGCTAT TCCTGTCAAT 360
ATCCATTCCA TGAATTTTAC CTGGCAAGCT GCAGGGCAGG CAGAATACTT CTATGAATTC 420
45 CTGTCTTGC GCTCCCTGGA TAAAGGCATC ATGGCAGATC CAACCGTCAA TGTCCCTCTG 480
CTGGGAACAG TGCCTCACA GGCATCAGTT GTTCAAGTTG GTTCCCATG TCTTGGAAAA 540
50 CAGGATGGG TGGCAGCATT TGAAGTGGAT GTGATTGTTA TGAATCTGA AGGCAACACC 600
ATTCTCCAAA CACCTCAAAA TGCTATCTTC TTTAAACAT GTCAACAAGC TGAGTGCCCA 660
GGCGGGTGCC GAAATGGAGG CTTTGTAAAT GAAAGACGCA TCTGCGAGTG TCCTGATGGG 720
55 TTCCACGGAC CTCACGTGA GAAAGCCCTT TGTACCCAC GATGTATGAA TGGTGGACTT 780
TGTGTGACTC CTGGMTCIG CATCTGCCCA CCTGGATTCT ATGGAGTGAA CTGTGACAAA 840
60 GCAAACTGCT CAACCACCTG CTTTAATGGA GGGACCTGTT TCTACCCTGG AAAATGTATT 900

	TSCCCTCCAG GACTAGAGGG AGAGCAGTGT GAAATCAGCA AATGCCACACA ACCCTGTGCA	960
	AATGGAGGTA AATGCATTGG TAAAAGCAAA TGTAAGTKTT CCAAAGGTTA CCAGGGAGAC	1020
5	CTCTGTTCAA AGCCTGTCTG CGAGCCTGGC TGTGGTGAC ATGGAACCTG CCATGAACCC	1080
	AACAAATGCC AATGTCAAGA AGGTTGGCAT GGAAGACACT GCAATAAAG GTACGAAGCC	1140
10	AGCCTCATAC ATGCCCTGAG GCCAGCAGGC GCCAGCTCA GGCAGCACAC GCCTTCACTT	1200
	AAAAAGGCCG AGGAGCGCGG GGATCCACCT GAATCCAATT ACATCTGGTG AACTCCGACA	1260
	TCTGAAACGT TTTAAGTTAC ACCAAGTTCA TAGCCTTTGT TAACCTTTCA TGTGTTGAAT	1320
15	GTTCAAATAA TGTTCATTAC ACTTAAGAAT ACTGGCCTGA ATTTTATTAG CTTCAATTATA	1380
	AATCACTGAG CTGATATTTA CTCTTCCTTT TAAGTTTCT AAGTACGTCT GTAGCATGAT	1440
20	.GGTATAGATT TTCTTGTTTC AGTGCTTTGG GACAGATTTT ATATTATGTC AATTGATCAG	1500
	GTTAAATTT TCAGTGTGTA GTTGGCAGAT ATTTTCAAAA TTACAATGCA TTTATGGTGT	1560
	CTGGGGCAG GGGAACATCA GAAAGGTTAA ATTGGGCAAA AATGCGTAAG TCACAAGAAT	1620
25	TTGGATGGTG CAGTTAATGT TGAAGTTACA GCAITTCAGA TTTTATTGTC AGATATTAG	1680
	ATGTTTGTTA CATTTTAAAA AATTGCTCTT AATTTTAAAA CTCTCAATAC AATATATTTT	1740
30	GACCTTACCA TTATTCAGA GATTCAGTAT TAAAAAATA AAAATTACAC TGTGGTAGTG	1800
	GCATTTAAAC AATATAATAT ATTCTAAACA CAATGAAATA GGAATATAA TGTATGAACT	1860
	TTTTCATTG GCTTGAAGCA ATATAATATA TTGTAAACAA AACACAGCTC TTACCTAATA	1920
35	AACATTTTAT ACTGTTTGTA TGTATAAAT AAAGGTGCTG CTTTAGTTTT CTGA	1974

40

(2) INFORMATION FOR SEQ ID NO: 204:

(i) SEQUENCE CHARACTERISTICS:

45

- (A) LENGTH: 1057 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

50

CGGCCTTCCG GGGCAACCGT TCGTCCAAC NCGGAAAGG GTCCTGGAGN CGGGAAC TAG	60
GAGCCTCGGA AGTCCAAGGG CGGAGCGCCC TTTGCTAATA AGCCAATCAG AACGTGAGAC	120
GCTCCGGTGG GNCGGTGCCG TCGAGCGCGG GGTGGAGTCT GGGTGACTTG GCTGGCGGGA	180
TCAAGTCAG CTGCTTCAGG CTGAGGTGGC AGATAGTGAG CGCTGGTGGC GGAGTTAAAG	240
TYAAAGCAGG AGAGTAATWA TGAATAGCGC AGCGGGATTTC TCACACCTAG ACCGTGCGGA	300

60

457

CGGGTTCTC AAGTTAGGGG AGAGTTTCGA GAAGCAGCCG CGCTGCGCTT CCACACTGTG 360
CGCTATGACT TCAAACCTGC TTCTATTGAC ACTTCTTCTG AAGGATACCT TGAGKTTGGC 420
5 GAAGKTGAAC AGKTGACCAT WACTCTGCCM AATATAGAAA GTTGAAGGAA GCAGTAAAT 480
TCAGTATCGT AAAGAACAAC AGCAACAACA ATGTGGAATT CASCAGGAC TCCCAATCTT 540
GTAAACATT CTCCATCTGA AGATAAGATG TCCCAGCAT CTCCAATAGA TGATATCGAA 600
10 AGAGAACTGA AGGCAGAAGC TAGTCTAATG GACCAGATGA GTAGTTGTGA TAGTTCATCA 660
GATTCCAAA GTTCATCATC TTCAAGTAGT GAGGATAGTT CTAGTGACTC AGAAGATGAA 720
15 GATTGCAAAT CCTCTACTTC TGATACAGGG NAATTGTGTC TCAGGACATC CTACCATGAC 780
ACAGTACAGG ATTCTTGATA TAGATGCCAG TCATAATAGA TTTCGAGACA ACAGTGGCCT 840
TCTGATGAAT ACTTTAAGAA ATGATTTGCA GCTGAGTGAA TCAGGAAGTG ACAGTGATGA 900
20 CTGAAGAAAT ATTTAGCTAT AAATAAAAT TTATACAGCA TGTATAATTT ATTTTGTATT 960
AACAATAAAA ATTCTAAGA CTGAGGGAAA TATGTCTTAA CTTTGTATGA TAAAGAAAT 1020
25 TAAATTTGAT TCAGAAAAA AAAAAAAAAA AACTCGA 1057

30 (2) INFORMATION FOR SEQ ID NO: 205:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 721 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

40 GAATTCGGCA CGAGTCATCC CTCTCCCTCT TCACTCCCT TACTCTTACT CTGTTTTTTG 60
TGCTCCAGAC AGACAGACCC TACCTCTTTT GCTTCTTTT TGTGTGTTG TTTTGAGATG 120
GAGTGTGCT CTGTGTGCC AGGCTGGAGT GCAGTGGCGC AATCTCGGCT CACCACAACC 180
45 TCTGCTCCC GGGTTCAAGC AATTCTCCTG CCTCAGCCTC CCGAGAAGCT GGGGATTACA 240
GGCATGCGCC ACCACACCCA GCTNAATTTT ATATTTTTAG TAGAGATGGT GTTCTCCAT 300
50 GTTGGTCAGG CTGGCCTCAA ACTCCAACC TCAGGTGATN CCGCTGCTT TGGCCTCCCC 360
AAAGTGCTGG GATTACAGGC GTGAGCCACT GCGCCAGCC TCTTTTGCTC CTTTATACTC 420
ATTAACCTAC GCCTGTAATC CCTGTTTTG GAGCCAAAG TGAGAAGGTT GCTTGAGGCC 480
55 AAGAGTTTGA GACTAGCCTG GGCAACACAG CAAGATGCCA TCTTTATAAT AAAAATAAAA 540
ATAAAAATCA ATTAGCTGG CATGGTGAA CGCACCTGTA GTCCAGCCA ATGAGAGGC 600
60 TGAAGTGGGA GGATCATGA GCCCAGGAGT TGAGGTTGCA GTGAGCCATG ATCATGTCAC 660

TACACTCAGC CTGGGCAATA GAGGACATG TTGTCTCTAA AAAAAAAAAA AAAAAACTCG 720

A 721

5

(2) INFORMATION FOR SEQ ID NO: 206:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2465 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

15

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

20

CCACCATTTA TCCAACTGAA GAGGAGTTAC AGGCAGTTCA GAAAATGTT TCTATTACTG 60

AACGTGCTTT AAAACTCGTT TCAGACAGTT TGTCTGAACA TGAGAAGAAC AAGAACAAAG 120

AGGGAGATGA TAAGAAAGAG GGAGGTAAAG ACAGAGCTTT GAAAGGAGTT TTGCCAGTGG 180

25

GAGTATTGGC AAAAGGATTA CTTCTCCGAG GAGATAGAAA TGTCACCTT GTTTTGCTGT 240

GCTCAGAGAA ACCTTCAAAG ACATTATTAA GCCGTATTGC AGAAAACCTA CCCAACAGC 300

30

TTGCTGTTAT AAGCCCTGAG AAGTATGACA TAAAATGTGC TGTATCTGAA GCGGCAATAA 360

TTTTGAATTC ATGTGTGAA CCCAAATGC AAGTCACTAT CACTGACA TCTCCAATTA 420

TTGAGAAGA GAACATGAGG GAAGGAGATG TAACCTCGG TATGGTGAAA GACCCACCGG 480

35

ACGTCTTGA CAGGCAAAAA TGCCTGACG CTCGGCTGC TCTACGCCAC GCTAAGTGGT 540

TCCAGGCTAG AGCTAATGGT CTGAGTCCT GTGTGATTAT CATACGCATT CTTGAGACC 600

40

TCTGTCAGCG AGTTCCAAC TGGTCTGATT TTCCAAGCTG GGCTATGGAG TTACTAGTAG 660

AGAAAGCAAT CAGCAGTGCT TCTAGCCCTC AGAGCCCTGG GGATGCACTG AGAAGAGTTT 720

TTGAATGCAT TTCTTCAGG ATTATTCTTA AAGGTAGTCC TGGACTTCTG GATCCTTGIG 780

45

AAAAGGATCC CTTTGATACC TTGGCAACAA TGA CTGACCA GCAGCGTGAA GACATCACAT 840

CCAGTGACA GTTTGCATTG AGACTCCTTG CATTCCGCCA GATACACAAA GTTCTAGGCA 900

50

TGGATCCATT ACCGCAAATG AGCCAACGTT TTAACATCCA CAACAACAGG AAACGAAGAA 960

GAGATAGTGA TGGAGTTGAT GGATTTGAAG CTGAGGGGAA AAAAGACAAA AAAGATTATG 1020

ATAACTTTTA AAAAGTGTCT GTAAATCTTC AGTGTAAAA AACAGATGC CCATTGTGTG 1080

55

GCTGTTTTTC ATTCAATAA ATGTCTACAT TGAAAAATTT ATCAAGAATT TAAAGGATTT 1140

CATGGAAGAA CCAAGTTTTT CTATGATATT AAAAAATGTA CAGTGTAGG TATTATTGA 1200

60

ATGGAAGAC ACCCAAAAAA AAAAATGTGC TCCGACTAGG GGGAAAACAG TAGTTCCGAT 1260

459

TTTTCCCAT TATTTTATT TTATTTCTG GTGCCCTAG CTCCCCCCC TATTTTGTG 1320
 TCTTTTATTA ACTAGTGCAT TGTCTTATTA AATCTTCACT GTATTTAATG CAGGATGTGT 1380
 5 GCTTCAGTTG CTCGTGTAT TTTGATAITTT TAATTTAGAG GTTTTGTGTG CTTTGTGACA 1440
 CTAGTTGTAA GTTACTTTGT TATAGATGGT ATCCTTTACC CCTTCTTAAT ATTTTACAGC 1500
 10 AGTACGTTTT TTTGTAACGT GAGACTGCAG AGTTTGTMTT TCTATATGTG AAGGATTACA 1560
 ACACAAAAG TTATCCTGCC ATTCGAGTGC TCAGAACTGA ATGTTTCTGC AGATCTGTG 1620
 GCATTTGTCT CTAGTGTGAT ATATAAGGT GTAATTAAGA CAGAGTTCTG TTAATCTAAT 1680
 15 CAAGTTTGCT GTTAGTTGTG CATTAGCAGT ATAAAAGCTA ATATATACTA TATGGTCTTG 1740
 CAACAGTTTT AAAGCCTCTG CATAATGAT AATAAAAATG CATGACATTC TTGTTTTTAA 1800
 TAGACTTTTA AAATCATAAT TTTAGGTTTA ACACGTAGAT CTTTGTACAG TTGACTTTTT 1860
 20 GACATAGCAA GGCCAAAAT AACTTTCTGA ATATTTTTTT CTGTGTATA AGTGGAAAGG 1920
 GCATTTTTCA CATATAAGTG GGCTAACCAA TATTTTCAA AGAACTTCAT CATGTACAA 1980
 25 CTAACAACAG TAACTAGCCC TTAATTATGG TGACAGTTCC TTATTGGTGT GTGTGAGATT 2040
 ACTCTAGCAA CTATTACAGT ATAACACAGA TGATCTTCTC CACACACCCC ATCACCAGCA 2100
 TAATTTACAG TTCTGTTAAC AGTGAGGTG ATAAAGTATT ACTGATAAAA AATTATCTAA 2160
 30 GGAAAAAAC AGAAAATTAT TTGGTGTGGC CATCTTACCT GCTTATGTCT CCTACACAAA 2220
 GCTAAATATT CTAGCAGTGA TGTAATGAAA AATTACATCT TACTGTTGAT ATATGTATGC 2280
 35 TCTGGTACAC AGATGTCATT TTGTGTACAC AGCACTACAG TGAAATACAC AAAAAATGAA 2340
 ATTCATATAA TGACTTAAAT GTAFTATATG TTAGAATTGA CAACATAAAC TACTTTTGCT 2400
 TTGAAATGAT GTATGCTTCA GTAAAATCAT ATTCAAATTT AAAAAAAAAA AAAAAAAAAA 2460
 40 CTCGA 2465

45

(2) INFORMATION FOR SEQ ID NO: 207:

(i) SEQUENCE CHARACTERISTICS:

50

- (A) LENGTH: 1480 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

55

GAATTCGGCA CGAGCTCAAG CTGGCAGGTG GTCGGGGGAG CGGCCGGAGA GGAGCTGCCG 60
 GGAGTTCTGT CCTGTCAGGA CATGACACCA GTGGCATATC ACGGCCATGG GGTCTCAGCA 120
 60 TTCCGCTGCT GCTGCCCCCT CCTCTGCAG GCGAAAGCAA GAAGATGACA GGGACGGTTT 180

460

5 GCTGGCTGAA CGAGAGCAGG AAGAAGCCAT TGCTCAGTTC CCATATGTGG AATTCACCGG 240
GAGAGATAGC ATCACCTGTC TCAOGTGCCA GGGGACAGGC TACATTCCAA CAGAGCAAGT 300
AAATGAGTTG GTGGCTTTGA TCCACACAG TGATCAGAGA TTGCGCCCTC AGCGAACTAA 360
GCAATATGTC CTCCTGTCCA TCCTGCTTTG TCTCCTGGCA TCTGGTTTGG TGGTTTCTTT 420
10 CCTGTTTCCG CATTCAGTCC TTGTGGATGA TGACGGCATC AAAGTGGTGA AAGTCACATT 480
TAATAAGCAA GACTCCCTTG TAATCTCAC CATCATGGCC ACCCTGAAAA TCAGGAACTC 540
CAACTTCTAC ACGGTGGCAG TGACCAGCCT GTCCAGCCAG ATTCACTACA TGAACACAGT 600
15 GGTGAATTTT ACCGGGAAGG CCGAGATGGG AGGACCGTTT TCCTATGTGT ACTTCTTCTG 660
CACGGTACCT GAGATCCTGG TGCACAACAT AGTGATCTTC ATGCGAACTT CAGTGAAGAT 720
20 TTCATACATT GGCCTCATGA CCCAGAGCTC CTTGGAGACA CATCACTATG TGGATTGTGG 780
AGGAAATCC ACAGCTATTT AACAACTGCT ATTGGTTCTT CCACACAGCG CCTGTAGAAG 840
AGAGCACAGC ATATGTTCCC AAGGCCTGAG TTCTGGACCT ACCCCCACTG GGTGTAAGCA 900
25 GAGGAGGAAT TGGTTCACTT AACTCCCAGC AACATCCTC CTGCCACTTA GGAGGAAACA 960
CCTCCCTATG GTACCATTTA TGTTTCTCAG AACCAGCAGA ATCAGTGCCT AGCCTGTGCC 1020
30 CAGCAAATAG TTGGCACTCA ATAAAGATT GCAGAATTTA ATACAGATCT TTTCAGCTGT 1080
TCTTAGGGCA TTATAAATGG AATCATAAC GTGGTTCTAG GTTATCAAAC CATGGAGTGA 1140
TGTTGAGCTA GGATTGTGAG TGACCTGCAG GCCATTATCA GTGCCTCATC TGTGCAGAAG 1200
35 TCGCAGCAGA GAGGGACCAT CCAAATACCT AAGAGAAAAC AGACCTAGTC AGGATATGAA 1260
TTTGTTTCAG CTGTTCCCAA AGGCCTGGGA GCTTTTGTAA AAGAAAGAAA AAAGTGTGTT 1320
40 GGCTTTTTTT TTTTITAGAA AGTTAGAATT GTTTTACCA AGAGTCTATG TGGGGCTTGA 1380
TTCACCCTTC ATCCATTGGC TGGAACATGG ATTGGGGATT TGATAGAAAA ATAAACCTTG 1440
45 CTTTGTATTC AAAAAAAAAA AAAAAWAAA AAAAAGTGA 1480

(2) INFORMATION FOR SEQ ID NO: 208:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 872 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

CAGTATTCC CTCAGTACTG TAAGCAAAAG TGGTATGTTT TTCTTTCTTT ATGTCTACTC

60

461

TGTCCTCTGT GGCCTTCTGG TGTACCCCTC TCTTCCTAGC CATTTCAGTCT CTCTAGTCAC 120
 CTCCTAGTA GCTAGTGCTC TCTAAGTTTT TATTTAATTA GAACAACTCC ATTTCATTT 180
 5 CAAGGTAGGT CAATGGGGGG AAAAGCCTCA TGATTTAAAC TGAAGTTAAC AACACAGCTT 240
 TTAAATGAA AACTCATACT CCAACTTCTA AAGTATATTT GAGCTGATTT GTTCCAAAA 300
 CAAAGATATG CTGTACCTAA AACTGCTAAA AAAAAATAT AAAGACAAGG ACTAGGTGAT 360
 10 TAAGGGGAGA GAAAAATCAT YTCCTTTCCA GGAAACCTTT GCTAAAATAA GCAAACTTG 420
 ANTCATGCT TCATGGAAAC TGACACAAAG AAAAGAACT GATGGATTGC ACAGGCCTTG 480
 15 TTATAGAAAT AGATCTATAA AAAGATCTGT CCACAGGAAA TATACACCTT CTCCTGGTTC 540
 TGAACCTCAA TGGGGATTG TCACCTAGGT CTCCATCTAT AGGAATACCT TCACATACCT 600
 ATCTATTCAT GCACATATTC TGAACACAGG TACATACAAA ATTACAACAA AGGAAAAAAA 660
 20 TTCTATTGAA CACTTAAAAA TAGAACAGG CCAGGCACGG TGGCTCATGC TGTAATCCCA 720
 ACAATTTGGG AGGCTGAGGC TGGTGGATCA CCTGAGGTCA GGAGTGTGAG ACCAGCTTGG 780
 25 CCAACATGGT GAAACCCCGT CACTACTAAA AATACAAAAA AAATTAGCCT GTGTGGTGGC 840
 ACACTCNTAC AATCCNGGCT GACTCGGGAA AN 872

30

•

(2) INFORMATION FOR SEQ ID NO: 209:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1779 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

AATGCCAAG ACTGCACAAA ATTACAGTGC TAATGTATAT GGTTCAGTT CACATAAGA 60
 CAAAAGCATC TGTATGAAA TGAGTAGTAA TATTGGGTGG TTGATTGTT CTTAGCAGAC 120
 45 TTGGCTTCAT WTTGGTCTTG AGATAAAATG GCCAGCATAA ATGCTGTTA TATTCAGTT 180
 TTCTAGGTG TGTGTGTGCA GGCCACAGCA GCATGCCCTT GGTGTAGTCA GTGCCGAAS 240
 50 GGGTCTGTC CTCTTGAGC CTGCTGCAG GGATGGTCTC CTTTAAAGC AGGTGTGTG 300
 CAGCATTCAG TACACTGAAG GTAAGCTAAA CCATCAACAT CTCGGTGT TTAAGATGTT 360
 ATTTTATTGG AACAACTGAC AAATGAGGGA TGTTAGCTTT GTGCAGAAT TCCCTGCATG 420
 55 TGTGATAACT GATCTTGTTT TATTTTTTGG CATTGCAACT GTGCATAGT TACAATTTCT 480
 GTTGKTCAT CACATTTAAA ATTGGRAGAG AACGCGCTTG AKGGATAGAG CGCCTTCAGK 540
 60 GTACTGTTTC TTATTAACCT TACTTTTTTT AAATCAACTT GCTATAGACT TTATATACAT 600

5 TTTGTAAAT ATAGTTCCTA GTGACATAGA AACGATGCGT AGTTTTCATT TACTAATTAC 660
 AAATGTTGAG GCCTAATTCT GAAAGTCCTC ATATTTAAAG GCTAGACAAC GTAATGAAAT 720
 TTTTAACTAT TTGTATGTCA TTTTGAAAGT GTACTGCTTT ATGGTAAAG TGTTTTTCAT 780
 TTGTTCAATG TTTTCATTAT TTGTGATCAT GTTGTCTTTC AATACAGGCA TAAACCTTCC 840
 10 ACTCTGAAC AAAGCAGCTG CTTTTTAAAA GCGGTAATG CTTCTTTACC TTTTATTCTT 900
 TTTGTAAATG AAGCTTTTCT TTAAGAATGT GACTTTAAAG TGTGTCTAT TGCATAAAAC 960
 AGTTGACACT CACTTATTGT AAAGTGAAGA TTGTTCTACT GCATGTGAAG TGGACCATGC 1020
 15 AGATTTCTGT ATGTCTCAG TATGCATCAC TAGATAATAA AGTCTTTTGT GAACAAGGCA 1080
 TTTGTAGCCA TTTTAAAAAG TTTTGTCTT CAGTGCTGGT AAGTCAGGTA AACCATAAAT 1140
 20 AGTTAAAGC AACCTTTTGT TTTTCTCTG AAAGTTTTTA ATTGAAAGTA TTATTAGTTA 1200
 AAGATGTAA CCTAGCCAAA ATTACCAGTT TATTAATAAT TAGGATCCTA ATTATTTCAA 1260
 AAAATCCTAC AAATATTGTC AGCTTTCAGT GTAGTGAGAT TATTCCTGTA GGTATGGGG 1320
 25 TATAATTCAG GATTAACTA ATGTTTCTGC TATTTCTCA CTTTCTCTT TGATGGTGG 1380
 GAAAGAGAAA AAGGAAAACG GGGCACAGGC CATTOGACGC CTTCTCCAAG GGTCTGATT 1440
 30 TGCTGAGACA CCAGCTTCAC CTTCTTAACA AGGCACCTAA TTACAACAAG CATGCACATT 1500
 TTGGTGCATT CAAGAATGGA AAATCAGAAT AGCAGCATG ATTCTTCTGG TGCAGCTCAG 1560
 TGGAAGATGA TGACAACCAG AAGACATGAG CTAAGGGTAA GGGACTGTTT TGAAGAACCT 1620
 35 TTCCATTTAG TGATCAAGAT ATGGAAGCTG ATTTCTGAAA ATGCTCAGTG TGTACTCTAA 1680
 TTATTTATGG TACCATTGTA ATTGTAACCT GCATTTTAGC AGTGCATGTT TCTAATTGAC 1740
 40 TTAAGGGAA ACTGAATAAA ATATGCCTCT TATTATCAA 1779

45 (2) INFORMATION FOR SEQ ID NO: 210:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 2110 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

55 GCGGCCGCTG CAGCCCGGAG CTGAGCTAGC CGTCCGAGCC GAGCCGTCCG AGCCGGGGAA 60
 GCCGGCGCGT GCTGCCGCTC GTGGCGGCCA GAGGAGAGGA GAGGCAGCAG CATGCCGAGT 120
 60 GTCTGTGCC GACGCCTTGG AAAGCGGTCC CTCCTGGGAG CCGGGGTGTT GGGACCCAGT 180

	GCCTCGGAGG GGCCTCGGCT GCCCCACCCT CGGAGCCACT GCTAGAAGGG GCCGCTCCCC	240
	AGCCTTTCAC CACCTCTGAT GACACCCCTT GCCAGGAGCA GCCCAAGGAA GTCTTAAGG	300
5	CTCCCAGCAC CTCGGGCCCT CAGCAGGTGG CCTTMMAGCC TGGGCAGAAG GTTTATGTGT	360
	GGTACGGGG TCAAGAGTGC ACAGGACTGG TGGWGCAGCA CAGCTGGATG GAGGGTCAGG	420
	TGACCGTCTG GCTGCTGGAG CAGAAGCTGC AGGTCTGCTG CAGGGTGGAG GAGGTGTGGC	480
10	TGGCAGAGCT GCAGGGCCCC TGTCCCCAGG CACCACCCCT GGAGCCCGGA GCCCAGGCCC	540
	TGGCCTACAG GCCCGTCTCC AGGAACATCG ATGTCCCAA GAGGAAGTCG GACGCATGA	600
15	AATGGATGAG ATGATGGCGG CCATGGTGCT GACGTCCCTG TCCTGCAGCC CTGTTGTACA	660
	GAGTCTTCCC GGGACCGAGG CCAACTTCTC TGCTTCCCGT GCGGCCCTGG ACCCATGGAA	720
	GGAGAGTGGT GACATCTCGG ACAGCGGCAN CAGCACTACC AGCGGTCACT GGAGTGGGAG	780
20	CAGTGGTGTC TCCACCCCTT CGCCCCCCA CCCCAGGCC AGCCCCAAGT ATTTGGGGGA	840
	TGCTTTTGGT TCTCCCCAAA CTGATCATGG CTTTGAGACC GATCCTGACC CTTTCTGCT	900
25	GGACGAACCA GCTCCACGAA AAAGAAAGAA CTCTGTGAAG GTGATGTACA AGTGCCCTGTG	960
	GCCAACTGT GGCAAAGTTC TGCCTCCAT TGTGGGCATC AAACGACACG TCAAAGCCCT	1020
	CCATCTGGGG GACACAGTGG ACTCTGATCA GTTCAAGCGG GAGGAGGATT TCTACTACAC	1080
30	AGAGGTGCAG CTGAAGGAGG AATCTGCTGC TGCTGCTGCT GCTGCTGCCG CAGACCCCCA	1140
	GTCCCTGGGA CTCCCACTC CGAGCCAGCT CCCACCCCA GCATGACTGG CCTGCCCTTG	1200
35	TCTGCTCTTC CACCACCTCT GCACAAAGCC CAGTCTCTCG GCCCAGAACA TCCTGGCCCG	1260
	GAGTCTTCCC TGCCCTCAGG GGCTCTCAGC AAGTCAGCTC CTGGGTCTTT CTGGCACATT	1320
	CAGGCAGATC ATGCATACCA GGCTCTGCCA TCCTTCCAGA TCCAGTCTC ACCACACATC	1380
40	TACACCACTG TCAGCTGGGC TGCTGCCCCC TCGCCGCTT GCTCTCTTTC TCGGTCCGG	1440
	AGCCGGTCCG TAAGCTTCAG CGAAGCCCA GCAGCCAGCA CCTGCGATGA AATCTCATCT	1500
45	GATCGTCACT TCTCCACCCC GGGCCAGAG TGGTGCCAGG AAAGCCCGAG GGGAGGCTAA	1560
	GAAGTGCCGC AAGTGTATGG CATCGAGCAC CGGGACCAGT GGTGCACGGC CTGCGGTGG	1620
	AAGAAGGCCT GCCAGCGCTT TCTGGACTGA GCTGTGCTGC AGGTCTTACT CTGTTCTTGG	1680
50	CCCTGCCGGC AGCCACTGAC AAGAGGCCAG TGTGTACCA GCCCTCAGCA GAAACCGAAA	1740
	GAGAAAGAAC GGAAACACGG AGTTTGGGCT CTGTTGGCTA AGGTGTAACA CTTAAAGCAA	1800
55	TTTCTCCCA TGTGCGAAC ATTTTATTTT TAAAAAATA GAAACAAAA TATTTTCCC	1860
	CCTAAATAG GAGAGAGCCA AAAGTACCA AGGCTATTCA GCAGTGAACC AGTGACCAA	1920
60	GAATTAATTA CCTCCGTTT CCCACATCCC CACTCTCTAG GGGATTAGCT TGTGCGTGT	1980

464

AAAAGAAGGA ACAGCTCGTT CTGCTTCCTG CTGAGTCGGT GAATTCCTTG CTTTCTAAAC 2040
TCTTCCAGAA AGGACTGTGA GCAAGATGAA TTTACTTTTC TTAAAAAAA AAAAAAAAAA 2100
5 AAAAACTCGA 2110

10 (2) INFORMATION FOR SEQ ID NO: 211:

(i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 938 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

20 GGCACAGGAA AAAAAAGAAA AAAGAAAAA GAAAAAGTT TTTGTACCCA CAGATTAGCA 60
TTTTCTTGAT GTTTGAAAAA AGTTTAAGCT ATGTCCTAAT TTAAAAATGA GCACAAACTA 120
CTTAACAGAT GTCTGTCCC TCTTCTCTTA CTAAATTAT CTTTATTTTC ACCATCACCT 180
25 CCCAGTGCCG AACACCTGAN CTCGTGTGTT TGTTGGTTGA TCCTGGGTG CCAAGTTCCT 240
ATTTGGTCAG TCCTGGCCT GTGGGGCGGT CTCAGGAAGT GGCATGCTCT TCAMGRAGGA 300
30 TCGTTCATYT CCAGTATAAC CAAATTGTTA ATAATAGTTG ATAATTCCCA GCTTTTACCA 360
GATGARITTT GACTTATTTT TCCTCCTTTG ACCTGTTCAG AGCTAACATA TCTCGGTCAG 420
TTCGGAGAGG GTGGGGGATT TGAGAATGTG AGGAGGAGTG GGGTTAGAAT GGGTTTGCCT 480
35 ATCTGGGCAA GGAAAGAGTT CCTAGTCGAT TGGGCACAAT GACAAAATGA TTCCATGGAT 540
AGAATCGTCC CATGTGCTG GAACACCTCA CGTGTGTGTA ACGCCTTAA TTCTTGCCAT 600
40 CCCTTCTCTG ATTCCCCACC TCCTGTAGT TTCCACAGGA TTTATCTCTC TGTACCCCG 660
TCCTCCAACCT CTACTCTGTC AGCCTCTCCT CCATCCCTTA CTTCCCTTCT AAATTCCAGG 720
AGATGACCTC ACTTTGCAAA GCAAATTGGA GCCACCAAAT TGTAGCTCTC CTCGGTGGAA 780
45 ACTGCATCTG TGCTCATCCC TGCACCTTCT TGCAGAAAGC CGCCCCCTCA GGCCAAGATG 840
AGTGCTGGC CCCCATGGGA GACTCAGACA CTTTGACCCC TTGTGACTTC AGCATCTCCC 900
50 TCTTTAAAGA TTCTCTCCCA ACATTCACTC GTGCTCGA 938

55 (2) INFORMATION FOR SEQ ID NO: 212:

(i) SEQUENCE CHARACTERISTICS:

- 60 (A) LENGTH: 1551 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 212:

5	AGGCTGGACT AAGCATAGAG AACCAGGAGA GAAAGAAAGA TTTAAGAGAC TGAGTAATAT	50
	TTTTTGACAG ATCATTTAAG AAACAGTA ATTTTTTTTT TCTCCAAAAG GGCATGGGT	120
	TTTTTTTTGT TTGTTTTTT CTCTATTGG CACTTCTAG GGATTGGTCT ATAAATTTTT	130
10	TGAAAGATCA TAGGATAAAT TTCTTTGTAG CAACCTCCTA TTTTAGTGT TATGTTAGGG	240
	GARCCCCARG TGTCCTGCT GATACGCCAT TAGGGCCACT TCTCAGCCTC TGGCTACATC	300
	ATAATGCTTT TTTTCTATC TTGCCAAAGT TTCCMGAAAA TTKAKGTTTT CTAATTTTAA	350
15	AAAAATGGT TGTCGAGATG GGATGGGACC TCTTTATAAG CCTGAAAAT AAGTGATTTN	420
	TTTTAAGTGC TATCTGCTA TAAACCTGAT TCTCACTTTT TTCTGTAGAC AACAGTTTTT	430
20	TATAATATAT CTATTTTGIG TGGACATTAT TTCTTTTAA CCAATACTGA AATTCATAG	540
	TGTAWACTTT CTCCACATTT TCTTTGATTA ATACTTYCTT AAAATAGACA CTTGGATTGG	630
25	CACCAGCTGT CACCAATAAA GCTGCCCTGA ACATTGTCAA TCAATCCTGT TAACCAATTT	650
	GAGAATTTTT CTGGAATGCT TAGTTAGGGA TGAAATGCT GGGTTATAGG TATGAGTATG	720
	CTTGATATAC TTTTCTCCAG AATGTCTACA CCTGTGTGTA CACCACATCT CCAGAGATAG	730
30	GGGAATCTTA TGTCCTGCT AACTGCTCTC GTTATTTAAT TTTCTGACAT TTGCCGCCGC	840
	CGCGCCCCC TGCCCCAAC ACACACATGG TATAAAGTGG TAGTTTCTTG TTTTAAATG	930
35	AACTTTTGAA TGATTTGAAT TTGGGCAITT CTTTGTATCC TGAGTTATTT TGGTTCCCG	950
	TTATGTGAAT ATCCTTTTCC TATGCTTTAA CTACTTTTCT AATTTGTCCC TTTTITNGGT	1020
	TATCAAATC CAGGCCATIG TCTATTCCAT CGTCACTTTT GGGTATTGGA AACATCTTTC	1030
40	CATTCTGTAG CCGTCTGTT GAACATAAAT CTGATTTTT ATGTAATCAG ATTTTCTCC	1140
	TTACGGTTAT GTTCTTGAA TTTTATTTAA GAAATCTTT TCTATCCTGA GACCACAAA	1200
45	ATGTCCCCAC CATTTTCTTC TGTTCATAG TTTGCTTG TATGTTAAT CCTTTAAGGC	1250
	ATGTGTAGTT CATTTTATAT GGTGTGAAAT AGTCTTATT CATTTATTCA ACACATATTG	1320
	GTGGAGTGCC TGCTGATGGT AGTACTCTTC AGAGTACTTT GTATATATTT GTGAACACAT	1330
50	ATTCTTGCCC TGAAGCTTA TGTGTCTNT CAAGGTAGAT CCNACTCGG TTCCACCTG	1440
	TTTCTTCAG CCTCAGGAT GAATTCACA ATTTTACACA TAGCACCAGT TAAGGAATAG	1500
55	GCTTTATTGG AGAAAAGGAA GGCTTATTAG ACCAGCATCA GCAAAAAAA A	1551

60 (2) INFORMATION FOR SEQ ID NO: 213:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 997 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 213:

10 AGAGAGTCCT CAACAGAACC TAATCATGCT GGCACCTTA CTTCTACTCT CTAGCCCTCA 50
 GAACTGAGAG AACATAAACT CCAGTTGTTT AAGCTACCCA GCTCTGGTA TTGTTATA 120
 TAGCCCAAGC TAAGTCAGGT GGAAAGGCAG AAATATTTTG AGAAGATCA TTCTACAA 180
 15 AACAGAGTTG TTCTAAATGA AATGGCCAGA TATTTCATCT TTTCTACTCT AGTATTTATG 240
 AAAGTTTCAT TAAACACCAC TTGCCAGCA CCCAGGCTG CTTCTCTAG AACGCCAAC 300
 20 AAAAGCAAAT GATTTGAGGA ACAAAGAGT GGACACAGAG CTTCTCAGAA GATGGCTCCA 360
 TCTTCTGAGA TGATCTTCTG AGATCATCA TTTTCTGAC CTGATGCTCT ACTCCAATG 420
 TAGTAGATAA GAGCAAAGAC ACTTCTGTAT CTTGTGGAAA ATGCTGGAGC CTTGTGATG 480
 25 GAGAGGCTGA CACTGGGACC AACAGAAGGC CGGACATTA TTTGCTGCAG CCTTCTGCA 540
 CCTGGGCCCT CTTCAGGCCT TGTACCTTGC ACTCCCATG CTTCTGTAGC AACTGGTAG 600
 30 CTGAAGTAG GTATTTGAAG AGATAATTTG CCCCCAACA AATTTCTT AAAGAAAA 660
 GGAAACCACT AATTCACCT TGACAAACCA GTTGTTCAG TTTTACTTT TGCAATTTG 720
 AAATTTCTC TTTGGACCA TATGATCTG TTACATTAGG GCTCATCAT GCTAAGATAC 780
 35 ACAGCTAGGT CTACCAGTG CCAGTGGTCA AGAATGAAG AACCTCTAG AGAGAGATCA 840
 GTTCTAATA ACCTAACAGT TTTCTTGGG TATTACAAA AAAAAAAAAA TTAGAATAA 900
 40 ATGTCAGTGC CATGCAGCA AGTACAGATA TGGAAATGA AGCTTGTCT ACAACTGCA 960
 GATTGTTTG TTAATAAAT TGATTGGGAT CACTCGA 997

(2) INFORMATION FOR SEQ ID NO: 214:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1496 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

GAATTCGGCA CGAGTGACCA CAGATATCTT TGGCTTTCAG CCTCACCACA ATGCTGTCCA 60
 CTATGTTTTT TTTAATGAT TGACATCTCA TGAATCCACA AATTTAGCGG CTTTCCATC 120

TTTTCCATCT TTGTCATAGC TTCATCACGC ACGATGGAGG TCACTTCAGC ACTATCCGGA 180
 GCGGCTCAC GGACAGATCR GTGAATTTCC TTTTCCTTTT TCTTGATGTA CCGGATTGTC 240
 5 GACTCGTTAA CATTGAGCTC ATGGCCAACA GCACTGTAAC TCATGCCTGA TTGGAGCTTA 300
 TCCAACACGC GGAMTTTCTC CGTAAGGSAM ATCAMGGTCT TCTTTCGCTT AGGAACACTG 360
 10 GGCARARCTT AARCACTACG CTTGGGGGCC ATTTTAGAAA GCAAAACCAC CCACAAAAG 420
 CAGAAAAAA AGTGTCACTA AACAGACTGN NGANAGGACT CTTTGTTTAC AGCAGAGGAG 480
 CTGCGACTAG AAGCGGGCGC TTCTCCCCAG TTCAAACCTC AGCTGGGAAC CTTACCTCCG 540
 15 CCAACTCAA ATTTTCAACC TCTGCGCATG CCCGGGAAAS AAACCCCGAG AACAGTACCG 600
 TGATGATTGA TTTTAGGGTT ACAAATACAT TTTAGCAAGT AAGTGAATTT GGCATTACGA 660
 ATTAATGATT AATGAAGGTC AACTGTATTT CCATAGATAT GTAAITTTAT TTAAGCAGGT 720
 20 TTATTATATT AAGCGGSSA GGCAGCGCCG AAGACTACAA GTTCCAGCAT GCACCGCGTC 780
 CGGGCGGGTT CCGGCTCCCA GCGAGGCTT CAGGGAAGCC AGCCCGGAGG CATCGGCGCG 840
 25 AAGTGTCTGA GGGCAACCAC GTAGTACTCT CTGCGCATGT GCAAAGCGCT GTCGGGGGCC 900
 GCCCTAGCTG CGTGCGCCG CGCCGGGGCT CTATGGTCTC TCCCTAGAGC TTTGCCGTTG 960
 GAGGCGGCTG CTGCGGTCTT GTGAGTTGA CCAGCGTCGA GCGGCAGCAA CATGGAGGAA 1020
 30 TTCGACTCCG AAGACTTCTC TACGTCGAG GAGGACGAGG ACTACGTGCC GTCGGGTGAG 1080
 CGATTCCGCC TGAGGCGAGA AGCGAATTGC CCCGCCAC GCCTCACGTG AGGCGCGCTC 1140
 35 TGCCCCCGCG GCGCTCTGCC CTGTGGCCCA GGTGGTCCAG GGGGGCTCCT GTTCTCGAGC 1200
 GTCCGCTCCC TCAGGCCCT CATCTCGGC CGCTCCGGCC CGAGGCGTGT GCGCGTGGCG 1260
 GTTCTGTGCT CCCCTCCGT TGGGCAGCTC CGGCCGCCG CCCCTCTTGC AGCGCGGGAA 1320
 40 CGGCACATGG ACACGGCCCC TTGTGCTAG GGAGCGCTGT CGGTGAGCCC CGAACGACAA 1380
 CGCTGCTTCA GAAGTCGGGG CGGCAGTTCG AGCCTTGAA GTTTTTCAT GCCCTGGCCC 1440
 45 GAGAGAGCTG CTGGCCAACA ACCCGTCAA GATAGAGCTG TCCGNTCTCC GNTGCG 1496

50 (2) INFORMATION FOR SEQ ID NO: 215:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1308 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

60 TTGGCANCNG GGAGAGGGAA AGAGGAGGAA ATGGGGTTTG AGGACCATGG CTTACCTTTC

60

CTGCCCTTGA CCCATCACAC CCCATTTCCT CCTCTTTCCC TCTCCCGCT GCCAAAAAA 120
 AAAAAAAGG AAACGTTTAT CATGAATCAA CAGGGTTTCA GTCCCTTATCA AAGAGAGATG 180
 5 TGGAAAGAGC TAAAGAAACC ACCCTTTGTT CCCAACTCCA CTTTACCCAT ATTTTATGCA 240
 ACACAAACAC TGTCCTTTTG GGTCCCTTTC TTACAGATGG ACCTCTTGAG AAGAATTATC 300
 10 GTATTCCACG TTTTGTAGCC TCAGGTTACC AAGATAAATA TATGTATATA TAACCTTTAT 360
 TATTGCTATA TCTTTGTGGA TAATACATTC AGGTGGTGCT GGGTGATTTA TTATAATCTG 420
 AACCTAGGTA TATCCTTTTG TCTTCACAG TCATGTTGAG GTGGGCTCCC TGGTATGGTA 480
 15 AAAAGCCAGG TATAATGTAA CTTACCCCA GCCTTTGTAC TAAGCTCTTG ATAGTGGATA 540
 TACTCTTTTA AGTTTAGCCC CAATATAGGG TAATGGAAT TTCTGCCCT CTGGGTTCCC 600
 20 CATTTTACT ATTAAGAAGA CAGTGATAA TTTAATAATG CCACCAACTC TGGCTTAGTT 660
 AAGTGAGAGT GTGAAGTGTG TGGCAAGAGA GCCTCACACC TCACTAGGTG CAGAGAGCCC 720
 AGGCCTTATG TTAAATCAT GCACTTGAAA AGCAAACCTT AATCTGCAA GACAGCAGCA 780
 25 AGCATTATAC GGTCACTCTG AATGATCCCT TTGAAATTTT TTTTGTGTTT GTTTGTTTAA 840
 ATCAAGCCTG AGGCTGGTGA ACAGTAGCTA CACACCCATA TTGTGTGTTT TGTAATGCT 900
 30 AGCTCTCTTG AATTTGGATA TTGGTTATTT TTTATAGAGT GTAAACCAAG TTTTATATTC 960
 TGCAATGCGA ACAGGTACCT ATCTGTTTCT AAATAAACT GTTACATTC ATTATGGGT 1020
 ATGTATGACC TTCATTTTCC AAGAAATAGA ACTCTAGCTT AGAATTATGG ATGCTCTAAA 1080
 35 ATGTCAGAAT GGGAACTCTC CTCGAAGTTC TCCCAAATC AGAGACAGCA CTGCCTTCTC 1140
 CTAAATGATT ATTCTTTTCT CCTGTTTTTC TGGTATTTTC TAGGCATCCT TCTACCACA 1200
 40 GCCATAACCC TTTTACTT CCATTAGGCC GTATAACTGG NGGGACNGCT GGTGGGTATA 1260
 TAATACTGGT WCCAACAMAG GGGTCTGGA TGTACACMAG GTTATCTT 1308

45

(2) INFORMATION FOR SEQ ID NO: 216:

50

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1705 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 216:

TGGCCATGGA AGCGCTAGAA GGTTAGATT TTGAAACAGC AAAGAAGGAT TTCCTTGGAT 60
 CTGGAGACCC CAAAGAAACA AAGATGCTAA TCACCAAACA GGCTGACTGG GCCAGAAATA 120

60

	TCAAGGAGCC CAAAGCCGCC GTGGAGATGT ACATCTCAGC AGGAGAGCAC GTCAAGGCCA	180
	TCGAGATCTG TGGTGACCAT GGCTGGGTTG ACATGTTGAT CGACATCGCC CGCAAACCTGG	240
5	ACAAGGCTGA GCGCGAGCCC CTGCTGCTGT GCGCTACCTA CCTCAAGAAG CTGGACAGCC	300
	CTGGCTATGC TGCTGAGACC TACCTGAAGA TGGGTGACCT CAAGTCCCTG GTGCAGCTGC	360
	AGTGGAGACC CAGCGCTGGG ATGAGGCCCTT TGCTTTGGGT GAGAAGCATC CTGAGTTTAA	420
10	GGATGACATC TACATGCCGT ATGCTCAGTG GCTAGCAGAG AACGATCGCT TTGAGGAAGC	480
	CCAGAAAGCG TTCCACAAGG CTGGGCGACA GAGAGAAGCG GTCCAGGTGC TGGAGCAGCT	540
15	CACAAACAAT GCCGTGGCGG AGAGCAGGTT TAATGATGCT GCCTATTATT ACTGGATGCT	600
	GTCCATGCAG TGCCTCGATA TAGCTCAAGA TCCTGCCAG AAGGACACAA TGCTTGGCAA	660
	GTTCTACCAC TTCCAGCGTT TGGCAGAGCT GTACCATGGT TACCATGCCA TCCATGCCA	720
20	CACGGAAGAT CGTTTCAGTG TCCATCGTCC TGAAACTCTT TTCAACATCT CCAGGTTCTT	780
	GCTGCACAGC CTGCCCAAGG ACACCCCTC GGGCATCTCT AAAGTGAAAA TACTCTTCAC	840
25	CTTGGCCAAG CAGAGCAAGG CCTTCGGTGC CTACAGGCTG GCCCGGCACG CCTATGACAA	900
	GCTGCGTGGC CTGTACATCC CTGCCAGATT CCAAAAGTCC ATTGAGCTGG GTACCCTGAC	960
	CATCGCGGCC AAGCCCTTCC ACGACAGTGA GGAGTTGGTG CCCTTGTGCT ACCGCTGCTC	1020
30	CACCAACAAC CGCTGCTCA ACAACCTGGG CAACTGTGCT ATCAACTGCC GCCAGCCCTT	1080
	CATCTTCTCC GCCTCTTCTT ACGACGTGCT ACACCTGGTT GAGTTCTACC TGGAGGAAGG	1140
35	GATCACTGAT GAAGAAGCCA TCTCCCTCAT CGACCTGGAG GTGCTGAGAC CCAAGCGGGA	1200
	TGACAGACAG CTAGAGATTT GCAAACAACA GCTCCAGAT TCTTGCGGCT AGTGGGAGAC	1260
	CAAGGGACTC CATCGGAGAT NAGGACCCGT TCACAGCTAA GCTRAGCTTT GAGCAAGGTG	1320
40	GCTCARAGTT CGTGCCAGTG GTGGTGAGCC GGCTGGTGCT GCGCTCCATG AGCCGCCGGG	1380
	ATGTCTCAT CAAGCGATGG CCCCACCCC TGAGGTGGCA ATACTTCCGC TCACTGCTGC	1440
45	CTGACGCCTC CATTACCATG TGCCCTCTCT GCTTCCAGAT GTTCCATTCT GAGGACTATG	1500
	AGTTGCTGGT GCTTCAGCAT GGCTGCTGCC CCTACTGCCG CAGGTGCAAG GATGACCTG	1560
	GCCCATGACC AGCATCTTGG GGACGGCTG CACCTCTGC CGCCTTGGG GTCTGCTGGG	1620
50	CTGTGAAGGA GAATAAAGAG TTAACTGTC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1680
	AAAAAAAAAA AAAAAAAAAA AAANA	1705
55		

(2) INFORMATION FOR SEQ ID NO: 217:

60

(i) SEQUENCE CHARACTERISTICS:

470

- (A) LENGTH: 999 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 217:

AGCAAATCAC CTTAACGATC TGGAATGAAA CTGTGACCAG TGCCGCCCTG GGTGGTTCCTG 60
10 GAGAGACTGC CGTCTTCTTG TTTGGCCATA GGTGCTGGGG CCCCGGCTTC AGTCACTGTC 120
TCAGACAGKA GTCCCGATAA GCAGATCACC AGTCCTCCAC TGTCTTCCT GTCCGGCCTTG 180
CTGCATGAGA AGATAGCTGC TTCTCCCTC TTTTCTTACA CTGTAAATTA TTGTTTAC 240
15 ATTGAGTGYC TTAATAATAG TYTACAAATA CTATGTATTT ATGCAAACT GTTAAAGTTC 300
TCATCTGTTA TGATTGGATA CTGGTCTTG TCAGTAGTGG TCAGCATTGG GTGTGAGCT 360
20 TGTCTACTC CATACGTGTT TATCTGCTA TGCATTTTAC ATTGTGTGTT CACATCTATT 420
CCAAGGAGCC TTGCTAGAAA CAACACTGGC GGTCTCTGCA GGCCAGGCAG GCATTGGCCC 480
ATGCTGTGTC CCATAGGAGC CAATGGAAAG AACGTAGCTT GGTCTGCTAG CCAGCCGTGG 540
25 GGTGGCGCAG GCCAGGCAGC CTCTGCACCA GAGTCCAGCA CCTGCCCAIT CCCAGTCAC 600
ACAATCATAC TCTTCTTTCA TAGAGATTTT ATTACCACCT AGACCACCT AGTTTCTCTC 660
30 TCTGTTAGTG TCTGAGCTC TTTTGCAACA AAATGTAGGT ACAGACCAAT CCCTGTCCCT 720
TCCCAATCA GGAGCTCCAC ACCATGAGTT GTTGGTTTT CCAGAAGCTG CCAGTGGGTT 780
CCCGTGAATT GCGTTAAGAT ATCGATGATK TTTTATTG TTTTCTTCT GTTTTTTTTA 840
35 AATAATATAT TTAAGGCAG TATCTTTGT ACTGTGAATT TGCAAGAGAA GATGCAGAAT 900
GCATTTTTT TTTACTTCTG TTGGTGTGTA TTGTATATAG TGTGTGCT TCTTGTGATG 960
40 AAAATAAACT TTTCTTTAT AAAAAAAAAA AAAAAAAC 999

45 (2) INFORMATION FOR SEQ ID NO: 218:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 941 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 218:

55 GGCAAGAGTA GCATTTCATT TAATCTGCAG GTATATTCTC CCAACAGTTT ATTGTCAATG 60
GATGTCCTCA GCCAAGATTG TRAGGCAGAG AGGAGCTGTC CCAACCTACT ATACCACCGA 120
GGCTGGAGAG ATCATATTTT TGGTATTAAA CTGGAGTCTC TCCATCCTTC ACATTGTTGA 180
60

471

TGTCTCTGT AGCAAACCGG AAAAGTCAGT GACAGAAGAT GCCGCTAGCG GTTTGAGCCA 240
 GAGAATGACA GCTCTGGTTT GGAGAAAAGG GCCCGATGGT GGCTCTAGAA AGCCCATCCT 300
 5 TCTGCTCTTC TTTTTCCTCC CCTTATATT GTGCTTTCAT TCATTCATTC ATTCATCAAA 360
 CATTTGTGTA GCACCTATTA TGTGTCAAGC TCTGTGCTAG CCTCTGGAAA ACCTGCCCTC 420
 ATGTAGCTCA CTGTGGAGTA GGAGAAACAA TGACTACACT ATGATAAGCA CGGGTTGTCA 480
 10 GGGTCTCACA GAGCAGTGGC CCTCATCCA GACCGATGAG GTCAAAGAAG GCATCCAGGC 540
 GAGGATGGTG TCAGAGCTAA CTGAAGAATG AGAGGGAGCT GCACCASCAG GGGTTGGAAC 600
 15 TGAAGGTGGC AGTGCCTGGA GTCTTGATTC CAGCAGAGG AGAGCAGTCT GTGAAAAGGC 660
 ACCAAGGGTG GGAGAGGGCA GAGCACATGG AGGAACCTCA GGTAGTCTG GATGGCSCCTG 720
 GGGCAAAGCT AGAGAGGTAA GAAGAATCTA CAAATGTTCC TCGAGTTACA TGAACCTCCA 780
 20 TCCCAATAAA CCCATGGAAC ACGAAAAATT TAAGTCAGAA GTGCATTAA GGCTGGTCCG 840
 AGTAGAATGA TTTTACAAC GAATTGATCA CAACCAGTTA CAGATGCTT TGTTCCTTCT 900
 25 CCACTCCAC TGCTTCACCT GACTAGCCTT TAAAAA A 941

30 (2) INFORMATION FOR SEQ ID NO: 219:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 575 base pairs
 (B) TYPE: nucleic acid
 35 (C) STRANDEDNESS: double.
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

40 TAAGTGAAT CCCCCGGGT TGCAGGAAT TCGCACGAG GCATTCTGAG AAGCTTAAGA 60
 CATACTTTGA AGACAACCT AGGGACCTCC AGCTGCTGCG GCATGACCTA CCTTGCACC 120
 CCGCAGTGGT GAAGCCCCAC CTGGGCCATG TTCTGACTA CCTGGTTCCT CTGCTCTCC 180
 45 GTGGCCTGGT RCGCCCTCAC AAGAAGCGGA AGAAGCTGTC TTCTCTTGT AGGAAGGCCA 240
 AGAGAGCAAA GTCCAGAAC CCACTGCGCA GCTTCAAGCA CAAAGGAAAG AAATTCAGAC 300
 50 CCACAGCCAA GCCCTCCTGA GGTGTGTGG CCTCTCTGGA GCTGAGCACA TTGTGGAGCA 360
 CAGGCTTACA CCCTTCGTGG ACAGGCGAGG CTCTGGTGCT TACTGCACAG CCTGAACAGA 420
 CAGTCTGGG GCGGCGAGT CTGGGCCCTT TAGCTCCTTG GCACTTCCAA GCTGGCATCT 480
 55 TGCCCTTGA CAACAGAATA AAAATTTTAG CTGCCCAAA AAAAAAAAAA AAAAAAAAAA 540
 CTCGAGGGGG GGCCCGTACC CAATCGCCC TATAA 575

60

(2) INFORMATION FOR SEQ ID NO: 220:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3018 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

	GCCAGCCTTA CAGGTTTAC GTGAAATGAA AGCCATTGGA ATAGAACCCT CGCTTGCAAC	60
15	ATATCACCAT ATTATTCGCC TGTTCGATCA ACCTGGAGAC CCTTTAAAGA GATCATCCTT	120
	CATCATTTAT GATATAATGA ATGAATTAAT GGGAAAGAGA TTTTCTCCAA AGGACCOGGA	180
20	TGATGATAAG TTTTTCAGT CAGCCATGAG CATATGCTCA TCTCTCAGAG ATCTAGAACT	240
	TGCCTACCAA GTACATGGCC TTTTAAAAAC CGGAGACAAC TGGAAATTC A TTGGACCTGA	300
	TCAACATCGT AATTCTATT ATTCCAAGTT CTTCGATTG ATTTGTCTAA TGGAAACAAAT	360
25	TGATGTACC TTGAAGTGGT ATGAGGACCT GATACCTTCA GCCTACTTTC CCCACTOCCA	420
	AACAATGATA CATCTCTCC AAGCATTGGA TGTGGCCAAT CGCTAGAAG TGATTCTTAA	480
30	AATTTGGGAA AGATAGTAAA GAATATGGTC ATACTTTCCG CAGTGACCTG AGAGAAGAGA	540
	TCCTGATGCT CATGGCAAGG GACAAGCACC CACCAGAGCT TCAGGTGGCA TTTGCTGACT	600
	GTGCTGCTGA TATCAAATCT GCGTATGAAA GCCAACCCAT CAGACAGACT GCTCAGGATT	660
35	GGCCAGCCAC CTCTCTCAAC TGTATAGCTA TCCTCTTTT AAGGGCTGGG AGAACTCAGG	720
	AAGCCTGGAA AATGTTGGG CTTCAGGA AGCATAATAA GATTCTTAGA AGTGAGTTGC	780
40	TGAATGAGCT TATGGACAGT GCAAAAGTGT CTAACAGCCC TTCCAGGCC ATTGAAGTAG	840
	TAGAGCTGGC AAGTGCCCTC AGCTTACCTA TTTGTGAGGG CCTCACCCAG AGAGTAATGA	900
	GTGATTTTGC AATCAACCAG GAACAAAAGG AAGCCCTAAG TAATCTAACT GCATTGACCA	960
45	GTGACAGTGA TACTGACAGC AGCAGTGACA GCGACAGTGA CACCAGTGAA GGCAAATGAA	1020
	AGTGGAGATT CAGGAGCAGC AATGGTCTCA CCATAGCTGC TGGAAATCACA CCTGAGAACT	1080
50	GAGATATACC AATATTTAAC ATTGTTACAA AGAAGAAAAG ATACAGATTT GGTGAATTTG	1140
	TTACTGTGAG GTACAGTCAG TACACAGCTG ACTTATGTAG ATTTAAGCTG CTAATATGCT	1200
	ACTTAACCAT CTATTAATGC ACCATTAAAG GCTTAGCATT TAAGTAGCAA CATTCGGTIT	1260
55	TTGAGACACA TGGTGAGGTC CATGGCTCTT GTCATCAGGA TAAGCCTGCA CACCTAGAGT	1320
	GTGCGTGAGC TGACCTCAG ATGCTGTGCT CGTGCGATTG CCTCTCCTG CTGCTGGACT	1380
60	TCTGCCCTTG TTGGCCTGAT GTGCTGCTGT GATGCTGGTC CTTCACTTTA GGTGTTTATG	1440

	CAGTTCTAAC ACAGTTGGGG TTGGGTCAAT AGTTTCCCAA TTTCAGGATA TTTGATGTC	1500
	AGAAATAACG CATCTTAGGA ATGACTAAAC AAGATAATGG CAGTTTAGGC TGCACAACTG	1560
5	GTAAATGAC TGTAGATAAA TGTGTAAAT AGTGTACACG TTTGTATTTT TGTTAATATA	1620
	CCCCGTGCCA TAGTTTCTA ACTTGAACAG CCATGAATGT TTCATGTCTC CCTTTTTTTT	1680
10	TTGTCTATAG CTGTTACCTA TTTTAGTGGT TGAAATGAGA GCTAGTGATG ACAGAAGGAT	1740
	GTGGAATGTC TTCTTGACAT CATGTGTAT TGCTGGTAAT CAAGTTGGTA ACGACTACTT	1800
	CTAGCAGCTC TTACCACTAT GACTTAAGTG GTCCTGGAAG GCAGTAAGTG GAGGTTTGCA	1860
15	GCAATCCTGC CTTTCATGAGG GCTTCTACCA CTGACCACTT TGCACGTACC TGGCTCCCAG	1920
	ATTTACTTAG GTACCCACG AGTCGTCCAC ATAAGCAGCT TCATCTTAC CTTGCCAGAG	1980
20	TTGACAATTA TGGGATACTC TAGTCTACTT ATACTTGTGT TCCCATCTGT CTGCCATCCT	2040
	CTGAAGGCCA GGACCCAGTC ATACATCCTT AGAAACCAA GTATGGTTTT TGTTTTCTCT	2100
	TGGAATGTCA GGTCTTAAGG CATTTAATG AGGGACAAA AAAAAAAAAA GCCGATATAG	2160
25	TAGCTAGCTA CTTAAGCATC CATGGGTATT GCTCCATATC AAAGCAGATT TGCAGGACAG	2220
	AAAGAGTAAA TTAGCCTTCA GTCTTGGTTT ACAGCTTCCA AGGAGAGCCT TGGSCACCTG	2280
30	AAATGTTAAC TCGGTCCCTT CCTGTCTCTA GTTCATCAGC ACCTGCAGAT GCCTGACTCT	2340
	TGTTAGCCTT ACTATTCAAT ACAGTCCCTA GATTCACGGT ATGCCTCTTC CTATCCAGGC	2400
	ACCTATTCTG AATCACCATG TTGCTCTGCA GCTAGAGTTG ATAGGAGAAA ATCCATTTGG	2460
35	GTAGATGGCC TATGAATTTG TAGTAGACTT TCAAAATGAG TGATTTGTTA GCTTGGTACT	2520
	TTTAAGTTTG TGGTACAGAT CCTQCAAACC CATACTCTGA GCAATTAACT GCCTTGAACA	2580
40	TAGAGAAAAA TTAAGGCCTC ACAGGATGAG TCTCCATTCT CTGTAAATGC TTATTTTATC	2640
	ATAGTCTTTA GCCTCTAACT ATGAGTAAAA TGTCTCTTC GCCCGGTGT GGTGACTCAC	2700
	ACCTGTAACC TCAGCACTTT GGGAGGCAGA GGTGGGAGGA TCACITAGGT CCAGGAGTTC	2760
45	GAGACTAGCC TGGGCAACAT AGTGAGACAC CGGATCTACA AAAAAATAAA AAGCCAGACT	2820
	GGTGGTATGT ATCTGTGTCC CAGCTAATTG GGAGGGTGAG ATGGGAGGAT TGTTTGAGCC	2880
50	TAGGAGAGGG AGGTTGCAGT GAGCCGTGAT CGCACCCTG CACTCCAGCC TGGGCAACAG	2940
	AGCAAGACCC TGTCTTGAG AAACCAGAAT TTTGGAAGAG CAAATGGGGC TGAGTGCACT	3000
	GGCTCATGCC TGTAATCC	3018
55		

(2) INFORMATION FOR SEQ ID NO: 221:

60 (i) SEQUENCE CHARACTERISTICS:

474

(A) LENGTH: 968 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

GGCACGAGGG CCGCGGGACA TCCACGGGGC GCGAGTGACA CGCGGGAGGG AGAGCAGTGT 60
10 TCTGCTGGAG CCGATGCCAA AAACCATGCA TTTCTTATTC AGATTCATTG TTTTCITTTA 120
TCTGTGGGGC CTTTTACTG CTCAGAGACA AAAGAAAGAG GAGAGCACCG AAGAAGTGAA 180
AATAGAAGTT TTGCATCGTC CAGAAACTG CTCTAAGACA AGCAAGAAGG GAGACCTACT 240
15 NAAATGCCCA TTATGACGGC TACCTGGCTA AAGACGGCTC GAAATCTAC TGCAGCCGGA 300
CACAAAATGA AGGCCACCCC AAATGGTTTG TTCTTGGTGT TGGGCAAGTC ATAAAAGGCC 360
20 TAGACATTGC TATGACAGAT ATGTGCCCTG GAGAAAAGCG AAAAGTAGTT ATACCCCTT 420
CATTTGCATA CGGAAAGGAA GGCTATGCAG AAGGCAAGAT TCCACCGAT GCTACATTGA 480
TTTTTGAGAT TGAACTTTAT GCTGTGACCA AAGGACCACG GAGCATTGAG ACATTTAAAC 540
25 AAATAGACAT GGACAATGAC AGGCAGCTCT CTAAAGCCGA GATAAACCTC TACTTGCAAA 600
GGGAATTGA AAAAGATGAG AAGCCACGTG ACAAGTCATA TCAGGATGCA GTTTTAGAAG 660
30 ATATTTTTAA GAAGAATGAC CATGATGGTG ATGGCTTCAT TTCTCCCAAG GAATACAATG 720
TATACCAACA CGATGAACTA TAGCATATTT GTATTTCTAC TTTTTTTTTT TAGCTATTTA 780
CTGTACTTTA TGTATWAAAC AAAGTCMCTT TTCTCCMAGT TGTATTGCT ATTTTTCCCC 840
35 TATGAGAAGA TATTTTGATC TCCCAATAC ATTGATTTTG GTATAATAAA TGTGAGGCTG 900
TTTTGCAAAAC TTAATAAAAA ATTTAAAAAA ACTGGAGGGG GGCCCGTACC CAANTCGCG 960
40 NATATGAT 968

45 (2) INFORMATION FOR SEQ ID NO: 222:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1404 base pairs
(B) TYPE: nucleic acid
50 (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

55 CGTTTTCCGG CCGTGCCTTT GTGGCCGTCC GGCTCCCTG ACATGCAGCC CTCTGGACCC 60
CGAGGTTGGA CCTTACTGTG ACACACCTAC CATGCGGACA CTCTTCAACC TCCTCTGGCT 120
TGCCCTGGCC TGCAGCCCTG TTCACACTAC CCTGTCAAAG TCAGATGCCA AAAAAGCCGC 180
60

	CTCAAAGACG CTGCTGGAGA AGAGTCAGTT TTCAGATAAG CCGGTGCAAG ACCGGGGTTT	240
	GGTGGTGACG GACCTCAAAG CTGAGAGTGT GGTTCCTTGAG CATCGCAGCT ACTGCTCGGC	300
5	AAAGGCCCGG GACAGACACT TTGCTGGGGA TGTACTGGGC TATGTCACCTC CATGGAACAG	360
	CCATGGCTAC GATGTCACCA AGGTCTTTGG GAGCAAGTTC ACACAGATCT CACCCGCTCTG	420
10	GCTGCAGCTG AAGAGACGTG GCGTGAGAT GTTTGAGGTC ACGGGCCTCC ACGACGTGGA	480
	CCAAGGGTGG ATGCGAGCTG TCAGGAAGCA TGCCAAGGGC CTGCACATAG TGCCTCGGCT	540
	CCTGTTTGAG GACTGGACTT ACGATGATTT CCGGAACGTC TTAGACAGTG AGGATGAGAT	600
15	AGAGGAGCTG AGCAAGACCG TGGTCCAGGT GGCAAGAAGC CAGCATTTTCG ATGGCTTCGT	660
	GGTGGAGGTC TGAACCAGC TGCTAAGCCA GAAGCGCGTG GGCCTCATCC ACATGCTCAC	720
20	CCACTTGCC GAGGCTCTGC ACCAGGCCCG GCTGCTGGCC CTCTGGTCA TCCCGCTGC	780
	CATCACCCCC GGGACCGACC AGCTGGGCAT GTTCACGCAC AAGGAGTTTG AGCAGCTGGC	840
	CCCCGTGCTG GATGGTTTCA GCCTCATGAC CTACGACTAC TCTACAGCGC ATCAGCCTGG	900
25	CCCTAATGCA CCCCTGTCTT GGGTTCGAGC CTGCGTCCAG GTCTTGACC CGAAGTCCAA	960
	GTGGCGAAGC AAAATCCTCC TGGGGCTCAA CTTCCTATGGT ATGGACTACG CGACCTCCAA	1020
30	GGATGCCCGT GAGCCTGTTG TCGGGGCCAG GTACATCCAG ACACTGAAGG ACCACAGGCC	1080
	CCGGATGGTG TGGACAGCC AGGYCTCAGA GCACTTCTTC GAGTACAAGA AGAGCCGCAG	1140
	TGGGAGGCAC GTCGTCTTCT ACCCAACCTT GAAGTCCCTG CAGGTGCGGC TGGAGCTGGC	1200
35	CCGGGAGCTG GCGGTGGGG TCTCTATCTG GGAGCTGGCC AGGGCCTGGA CTACTTCTAC	1260
	GACCTGCTCT AGGTGGGCAT TGGGGCCTCC GCGGTGGACG TGTCTTTTC TAAGCCATGG	1320
40	AGTGAGTGAG CAGGTGTGAA ATACAGGCCT NCACTCCGTT TGCTGTGAAA AAAAAAAAAA	1380
	AAAAAAAAAA AAAAAAAAAA AAAA	1404

45

(2) INFORMATION FOR SEQ ID NO: 223:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 707 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

50

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

	NGCGCGCCTG CAGTCGACAC TAGTGGATCC AAAGAATTCG GCACGAGGGC AGGTCCAGGG	60
	CTCAGAAATC AGCTCTATTG ACGAATTCCTG CCGCAAGTTC CGCCTGGACT GCCCGCTGGC	120
60	CATGGAGCGG ATCAAGGAGG ACCGGCCCAT CACCATCAAG GACGACAAGG GCAACCTCAA	180

CCGCTGCATC GCAGACGTGG TCTCGCTCTT CATCACGGTC ATGGACAGSC TGGGCTTGA 240
 GATCCGCGCC ATGGATGAGA TCCAGCCCGA CCTGCSAGAG CTGATGAGAA CCATGCACCG 300
 5 CATGAGCCAC CTCCACCCCG ACTTTGAGGG CCGCCAGAGG GTGAGCCAST GGCTSCAGAC 360
 CCTGAGCGGC ATGTGGCGGT CAGATGAGCT GGCAGACTCA CAGGTGGSTC AGATCTGTT 420
 10 CGACCTGGAG TCAGCTACA ACGCTTCAA CCGCTTCTG CATGCCGTG CCGGGGGCAC 480
 TAGCCCTTGC ACAGAAGGGC AGAGTCAGG GCGATGGCTC CTGGTCTCTT GTCCGCCACA 540
 CAGGCCGTGG TCATCCACAC AACTCACTGT CTGCAGCTGC CTGTCTGCTG TCTGTCTTTC 600
 15 GTGTGAGAAC TTTTGGGCGG GGGCCCTCCC CCAATAAAG ATGCTCTCCG ACCTTCAAAA 660
 AAAAAAAAAA AAAAATCRG GGGGGGCGG GTCCCATCC CCCCCTT 707
 20

(2) INFORMATION FOR SEQ ID NO: 224:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1384 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 30
 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

GGGGAAGTC AGTGACAGCA GGAGTAAGG TGGGAGGCAG GACAGAGTG GGACACAGGT 60
 35 ATGGAGAGGG GGTTCAGCGA GCCTAGAGAG GGCAGACTAT CAGGGTGGCG GCGGTGAGAA 120
 TCCAGGGAGA GGAGCGGAAA CAGAGAGGG GCGAGAGACC GGGGCTCTTG TGGGTTCAG 180
 AGCCCTCAG CCATGTTGGG AGCCAAGCCA CAGTGGCTAC CAGGTCTCTT ACACAGTCCC 240
 40 GGGCTGCCCT TGGTTCGGT GCTTCTGGCC CTGGGGGCGG GGTGGGCTCA GGAGGGGTCA 300
 GAGCCCGTCC TGCTGGAGGG GAGTGGCTG GTGGTCTGTG AGCTTGGCG AGTCTCTGCA 360
 45 GGGGGGCTCG GGGAGCAGC CCTGGGAGAG GCACCCCTG GCGAGTGGC ACTTGTCTGG 420
 GTCCGAAGCC ATCACCATGA GCGAGCAGG GAAACCGCA ATGGCATAT TGGGGCCATC 480
 TACTTCGACC AGGTCTGGT GAGAGAGGG GGTGGCTTTG ACCGGGCTTC TGGCTCTTC 540
 50 GTAGCCCTG TCCGGGTGT CTACAGCTTC CCGTTCTATG TGGTGAAGT GTACAACCGC 600
 CAACTGTCC AGGTGAGCCT GATGCTGAC ACGTGGCTG TCATCTGAGC CTTTGCCAAT 660
 55 GATCTGAGC TGACCCGGA GGCAGCCACC AGCTCTGTG TACTGCCCTT GGACCTTGG 720
 GACCGAGTGT CTCTGCGCT GCGCGGGGG AATCTACTGG GTGGTGGAA AACTCAAGT 780
 TTCTCTGGCT TCTCATCTT CCTCTCTGA GGACCCAAGT YTTCTAGCA CAGAAATCCA 840
 60

477

GGGCTGACA ACTTCTCTCT GGGCTCTCTT GGGCCAGAAA CAGCAGAGGC AGGAGAGAGA 900
 CTCCCTCTGG YTCCTATCCC ACYTCTTTGC ATGGGAMCCT GTGCCAAACA CCCAAGTTTA 960
 5 AGARAARARY ARARCTGWWG CAGGTATACA GAGCTGGAAG TGGACCATGG AAAACATSGA 1020
 TAACCATGCA TCYTCTTGCT TGGCCACCTC CTGAAACTGT CCACCTTTGA AGTTTGAAC 1080
 10 TTAGTCCCTC CAMACTCTGA CTGCTGCCTC CTTCCTCCCA GCTCTCTCAC TGAGTTATYT 1140
 TCACTGTACC TGTTCCAGCA TATCCCCACT ATCTCTCTTT CTCTGTATCT GTGCTGTCTT 1200
 ATTCTCCTCC TAGGCTTCC TATTACCTGG GATTCCATGA TTCATTCTT CAGACCTCT 1260
 15 CCTGCCAGTA TGCTAAACCC TCCCTCTCTC TTTCTTATCC CGCTGTCCCA TTGGCCCAGC 1320
 CTGGATGAAT CTATCAATAA AACAACTAGA GAATGGTGGT CAAAAAAAAA AAAAAAAAAAC 1380
 20 TCGA 1384

(2) INFORMATION FOR SEQ ID NO: 225:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 760 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

GGGTCGACCC ACGCGTCCGC TGACCAGTCC GTTATAGATA CTTCTTCTTA TACCAAAAC 60
 35 GTTTAAACAG GTGCCACCAC AAGGGATGTC GTCTTACTC TCTGCGGGTC TTCAAGCATC 120
 CCTTTGTGGG AAARGTCTCT GGGCAAGCAC GTGGTATTTG GTCTGCTGCT TGCTTCCCTT 180
 40 TTTCCACCAG GGATGTGTGT ATCATAAGTC AAAACAACAG TATATTCCAA ATCTCAAAAG 240
 CTATTGTGGC CTGAGCACAA TTGAAATCTA GCAGAGTTTT TCCTATGTAG CTTTAGAGTA 300
 45 ACTCTTCTGC TTCTCTGTCA CTTACAATTC AGGTTCCTGCC TTTCCTAAG AGCATGAGCA 360
 GAAGAGTCTT CATGTGACGC TTAGTTCTAT TGCAGTCTTG GGTGAAACTA TTAAAGCWAT 420
 GGGGCTGCTK CTCCCANWT CCTCCCTAAC AATTCTGTGT GTGGACTTCT CATCTAAAAG 480
 50 GTTAGTGGCT TTGCTTGGG ATCAGTGCTC TCTATTGATG TTCTTGCTGG TCTCCAGACA 540
 CATCTCTGTT GCATTAGAC TTGAAAGACT TGTAGATGTG TGAATTCAG GCACAGGATG 600
 CTGAAAGCTA TGTTACTATT CTTAGTTTGT AAATGTCTCT TTTGATACCA TCATCTTGTT 660
 55 TTCTTTTGT AGGTATAAAT AAAAACAATG TTGACAATAA AAAAAAAAAA AAAAAAAAAA 720
 AAAAAAAAAA AAAAAAAAAA NAAAAAAAAA AAAAAAAAAA 760

60

(2) INFORMATION FOR SEQ ID NO: 226:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2057 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

CCGAGCCGGC TCGCCCGGGG GAATCCGTGC GGGGCGCTTC CGTCCCRGTC CCATCCTCGC 60
 15 CGCGCTCCAG CACCTCTGAA GTTTTGCAGC GCCCAGAAAG GAGGCGAGGA AGGAGGGAGT 120
 GTGTGAGAGG AGGGAGCAAA AAGCTCACCC TAAACATTT ATTCAAGGA GAAAAGAAAA 180
 AGGGGGGGCG CAAAATGGC TGGGGCAATT ATAGAAAACA TGAGCACCAA GAAGCTGTGC 240
 20 ATGTGTGGTG GGATTCTGCT CGTGTTCCAA ATCATCGCCT TTCGTGGTGGG AGGCTTGATT 300
 GCTCCAGGGC CCACAAOGGC AGTGTCTTAC ATGTCCGTGA AATGTGTGGA TGCCCGTAAG 360
 25 AACCATCACA AGACAAAATG GTTCGTGCCT TGGGGACCCA ATCATTGTGA CAAGATCCGA 420
 GACATTGAAG AGGCAATTCC AAGGGAAATT GAAGCCAATG ACATCGTGTT TTCTGTTTAC 480
 ATTCOCCTCC CCCACATGGA GATGAGTCTT TGGTTCCAAT TCATGTTGTT TATCCTGCAG 540
 30 CTGGACATTG CCTTCAAGCT AAACAACCAA ATCAGRGAAG ATGCAGAAGT CTCCATGGAC 600
 GTTTCOCCTGG CTACCGTGA TGACGCGTTT GCTGAGTGGG CTGAAATGGC CCATGAAAGA 660
 35 GTACCACGGA AACTCAAATG CACCTTCACA TCTCCAAGA CTCCAGAGCA TGGAGGGCCG 720
 GTTACTATGA ATGTGATGTC CTTCTTTTCA TGGAAATTGG GTCTGTGGCC CATGAAGTTT 780
 TACCTTTTAA ACATCCGGCT GCCTGTGAAT GAGAAGAAGA AAATCAATGT GGGAAATGGG 840
 40 GAGATAAAGG ATATCCGGTT GGTGGGGATC CACCAAAATG GAGGCTTCAC CAAGTGTGG 900
 TTTGCCATGA AGACCTTCTT TACGCCAGC ATCTTCATCA TTATGGTGTG GTATTGGAGG 960
 45 AGGATCACCA TGATGTCCCG ACCCCAGTG CTTCTGGAAA AAGTCATCTT TGCCCTTGGG 1020
 ATTTCCATGA CCTTTATCAA TATCCAGTG GAATGGTTTT CCATCGGGTT TGAATGGACC 1080
 50 TGGATGCTGC TGTTTGGTGA CATCCGACAG GCATCTTCTA TGCRATGCTT CTCTCCTTCT 1140
 GGATCATCTT CTGTGGGAG CACATGATGG ATCAGCACGA GCGGAACCAC ATCGCAGGGT 1200
 ATTGGAAGCA AGTCGGACCC ATTGCCGTTG GTCTTCTGTC CTCTTCATAT TTGACATGTG 1260
 55 TGAGAGAGGG GTACAACTCA CGAATCCCTT CTACAGTATC TGGACTACAG ACATTGGGAA 1320
 CAGAGCTGGC CATGGCTTTC ATCATCGTGG CTGGAATCTG CCTCTGCCTC TAACTTCTTG 1380
 60 TTTCTATGCT TCATGGTATT TCAGGTGTTT CGGAACATCA GTGGGAAGCA GTCCAGCCTG 1440

CCAGCTATGA GCAAAATCTG GCGGCTACAC TATGAGGGGC TAATTTTTAG GTTCAAGTTC 1500
 CTCAGGCTTA TCAAGCTGGG CTGCGCTGGC ATGAGTGTCA TCTTCTTCAT CGTTAGTCAG 1560
 5 GTACGGAAG GGTATTGGGA AATGGGGCGG CGTACAGTC CCAAGTGAAC AGTGCCTTTT 1620
 TCAAGGGCAT CTAAGGGATG TGGAACTGT ATGTCTTTCG TGTGATGTTC TTGTATGCAC 1680
 CATCCCATTA AACTATGGA GAGACCGGT CCAATGGAT GCAACTCCCA TGTAAATCGA 1740
 10 GGGAGATTG TGGTTGTGT GTTCGGAC TTTATCAGA ACTGTTCAGC GCTTCGAAAT 1800
 ATTCCTTCAT CAATGACAG GAGGTTCTG GTATTGAGT CACAAAGGCA ACACATGTTT 1860
 15 ATCAGCTTTC CATTTGAGT TTTACAGTC AGATTGATG TACTTGATA CGCACACAAA 1920
 TACACTCAT TACCTTTAT GTTAAATGT TAAATATAG GAAAAAAGCG TCAACAATAA 1980
 ATATTCTTTC AGTATGTGT TACTCTCTT AAAAAAAAA AAAAAAACTC GTGCCGAATT 2040
 20 CGCCAGAGC GGCACGA 2057

25

(2) INFORMATION FOR SEQ ID NO: 227:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2084 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

35 GGCAGAGGGC CATTTCTTGC AAGAGGCGAA ACCGCGATTC CTCTGTGCC CTCTCTCCC 60
 ACCAGTGTCT TTATAAAAT AGCTCTTGT ACCGGAATA ACTGTTCATT TTCACTCTCT 120
 40 CCGTCCTAGG TCACCTTTT CAGAAAAGA ATCTGCTCC TGGAAACCAG AAGAAAATA 180
 TGAGACGGGG AATCATCTG TATGTGTGT SCTGCTTTC GCTGAGTGTG TGGAGTCTTG 240
 CTCAGGTGTT AGTACAGTG TGTGTGATG TGTGGCTTG AGGGGAACCG CTTGTTCAGA 300
 45 GCTGTACTTG CCGCTCACT GAGAGAGAC TGCCCTTGGC TGCTCGTAGC GCGGGGCTT 360
 CTCTCTCTGT CATCATCCAG ACGAGCCAT GTCCGGGAGG CAGAAGTAC GGGGCGAGCT 420
 50 ACTGGAGGAG TGTGGGGCC TCCCTGGCT GCGGCTCCG CCGTGGGGCC CTGTGTCTGC 480
 TGACCACTTA TTTCTACTAC TCCCTCCAA ATGCGGTCCG CCGGCTTTC ACTTGGATGC 540
 TTGCGCTCTT GGGCTTCTC GAGGGCACTG AACATCTCC TGGGCTCAA GGGCTGGCC 600
 55 CCGCTGAGA TCTCTCTGT GTGTGAAAA GGAATTTCA ACGTGGCCA TGGGCTGGCA 660
 TGGTCATATT ACATCGATA TGTGGGCTG ATCTGCCCAG AGCTCCAGGC CCGGATTCGA 720
 60 ACTTACAATC AGCATTACAA CAACCTGCTA CCGGGTGCAG TGAGCCAGCG GTGTNATATT 780

480

CTCTCCCAT TGGACTGTGG GGTGCCTGAT AACCTGAGTA TGGCTGACCC CAACATTGCG 840
 TTCTTGATA AACTGCCCA GCAGACCGGT GACCGTGCTG GCATCAAGGA TCGGGTTTAC 900
 5 AGCAACAGCA TCTATGAGCT TCTGGAGAAC GGGCAGCGGG CGGGCACCTG TGTCTGGAG 960
 TACGCCACCC CCTTGCAGAC TTTGTTTGCC ATGTCACAAT ACAGTCAAGC TGGCTTTAGC 1020
 10 GGGGAGGATA GGCTTGAGCA GGCCAAACTC TTCTGCCGGA CACTTGAGGA CATCCTGGCA 1080
 GATGCCCTTG AGTCTCAGAA CAACTGCCGC CTCATTGCCT ACCAGGAACC TGCAGATGAC 1140
 AGCAGCTTCT CGCTGTCCCA GGAGGTCTCT CGGCACCTGC GGCAGGAGGA AAAGGAAGAG 1200
 15 GTTACTGTGG GCAGCTTGAA GACCTCAGCG GTGCCAGTA CCTCCACGAT GTCCCAAGAG 1260
 CCTGAGCTCC TCATCAGTGG AATGGAAG CCCTCCCTC TCCGCACGGA TTTCTCTTGA 1320
 20 GACCCAGGGT CACCAGGCCA GAGCCTCCAG TGGTCTCCAA GCCTCTGGAC TGGGGGCTCT 1380
 CTTCACTGGC TGAATGTCCA GCAGAGCTAT TTCCTTCCAC AGGGGGCCTT GCAGGAAGG 1440
 GTCCAGGACT TGACATCTTA AGATGCGTCT TGTCCCTTG GCCAGTCAT TTCCCTCTC 1500
 25 TGAGCCTCGG TGTCTTCAAC CTGTGAAATG GGATCATAAT CACTGCCTTA CCTCCCTCAC 1560
 GGTGTGTGTG AGGACTGAGT GTGTGGAAGT TTTTCATAAA CTTTGGATGC TAGTGTACTT 1620
 30 AGGGGGTGTG CCAGGTGTCT TTCATGGGGC CTTCCAGACC CACTCCCCAC CCTTCTCCCC 1680
 TTCTTTGCC CGGGGACGCC GAACTCTCTC AATGSTATCA ACAGGCTCCT TCGCCCTCTG 1740
 GCTCCTGGTC ATGTTCCATT ATTGGGGAGC CCCAGCAGAA GAATGGAGAG GAGGAGGAGG 1800
 35 CTGAGTTTGG GGTATGAAT CCCCCGCTC CCACCTGCA GCATCAAGGT TGCTATGGAC 1860
 TCTCTGCCG GGCAACTCTT GCGTAATCAT GACTATCTCT AGGATTCTGG CACCACTTCC 1920
 40 TTCCCTGGCC CCTTAAGCCT AGCTGTGTAT CGGCACCCCC ACCCCACTAG AGTACTCCCT 1980
 CTCACTTGGG GTTTCCTTAT ACTCCACCCC TTTCTCAACG GTCCTTTTTT AAAGCACATC 2040
 45 TCAGATTAAA AAAAAAAAAA AAAAAAAAAA AGGGGGGCN GCNT 2084

(2) INFORMATION FOR SEQ ID NO: 228:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2143 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

TCGACCCACG CGTCCGGTTG AATTCCCTGA CCTGCAAACA CATATTTATT AGCCTGACTC

60

	AAACAATGAA GCTATTAAAA CTTCGGAGGA ACATTGTAAA ACTCTCTTTG TATCGGCATT	120
	TCACCAACAC GCTTATTTTG GCAGTGGCAG CATCCATTGT GTTATCATC TGGACAACCA	180
5	TGAAGTTCAG AATAGTGACA TGTCAGTCGG ACTGGCGGGA GCTGTGGGTA GACGATGCCA	240
	TCTGGCGCTT GCTGTCTCC ATGATCCTCT TTGTCATCAT GGTCTCTGG CGACCATCTG	300
10	CAAACAACCA GAGGTTTGCC TTTTCACCAT TGTCTGAGGA AGAGGAGGAG GATGAACAAA	360
	AGGAGCCTAT GCTGAAAGAA AGCTTTGAAG GAATGAAAAT GAGAAGTACC AAACAAGAAC	420
	CCAATGGAAA TAGTAAAGTT AACAAAGCAC AGGAAGATGA TTTGAAGTGG GTAGAAGAGA	480
15	ATGTCCTTC TTCTGTGACA GATGTAGCAC TTCCAGCCCT TCTGGATTCA GATGAGGAAC	540
	GAATGATCAC ACACTTTGAA AGGTCCAAAA TGGAGTAAGG AATGGGAAGA TTTGCAGTTA	600
20	AAGATGGCTA CCATCAGGGA AGAGATCAGC ATCTGTGTCA GTCTTCTGTA CGGCTCCATG	660
	GGATTAAAGG AAGCAATGAC ATCCTGATCT GTTCTTGAT CTTTGGGCAT TGGAGTTGGC	720
	GAGAGGTGTC AGAACAAGA GAACATCTTA CTGAAAACAA GTTCATAAGA TGAGAAAAAT	780
25	CTACGAGCTT CTTATTTACA ACACGTCTGC CCCCTTTCCT CCCAGACTCT GACATGGATG	840
	TTCATGCAAC TTAAGTGTGT TGTTCCTGAA CTTTCTGTAA TGTTCATTT TTTAAATCTG	900
30	ACAAACTAAA AAGTTTAAAG TCTTCTAAAA GATTGTATC AACACCATAA TATGTAATCT	960
	CCAGGAGCAA CTGCCTGTAA TTTTATTATA TTTAGGGAGT TACATAGGTG ATGGGGGAAA	1020
	TGTTAACCTA CCTTTCATTT TCCCTGGGAAG TCAAGGTAC ATCTTGCAGA GGTGTGTTTG	1080
35	AGAAAAAAGG GCCCTTCTGA GTTAAGGAGC CATAGTTCTA TCAATGATCA AAAGAAAAAA	1140
	AAAAAAAAGA GAACTGTTA CAGTATGATT CAGATCATTT AAAAAAGCAA AATCAAGTGC	1200
40	AATTTTGT TT ACAAATGGTG TATATTAAAG ATTTTCTAT TTCAGATGTA CTTTAAAGAG	1260
	AAATATTAGC TTAACCTTT TGACATCTGC TATTGTGACA CATCCCATG CTGGCAATGT	1320
	GGTGCACT COGAACTTT TAACTACTGT TTTGTAAGCC TCCAAGGGTG GCATTGCAGG	1380
45	GTCCCTAGGC AATGTTTTGT TTGCCTTTAT GCAGAGAGGT GCTCCAAGTG CTGTGATTGA	1440
	GCACCGTGCT AGAGGAACTG TAATGCTTCA GAAGTTGTAG CTTATACAAA GGAAACAGGT	1500
50	CTTGCTGGCT TAATTAAAC AGTTATGCA TGAAGTAGCG TGGAGGCCCT GGACTGCTGC	1560
	TGTTCTTTA GGATGGACTG TTCTGGTATC TGGTATTGGT TTAGAGACTG TTAATAAGGG	1620
	ACATCACAAG GTGATGGGAT TCATTTGAAG CACTCTATTT CTGTTTTAAT GGTTTTATCC	1680
55	AATTTTGCCT TCCCAAGATT TTGTTCTAC ATAAAAAGTT CATGCCACTT TTTAATATAA	1740
	AAAAATTTAA CAAATTAAT GTATTTTCT CATTTTTTC AAACTTTTT TAAAGACTCT	1800
60	TTCTGTCAAA CTCATGAAAA ATTTCTTCT ATGGCTTTTA TTCTAGATTG TCTTATTTTC	1860

482

TGTTAAAACC AATGACCACA TGACCACAAT CTTCACTAAC TCATACTGCA GTGAAAGTGT 1920
TAACCCCTTAG GTAGTTTCTC TACAACCTCT TGCTATGGTG ATTTTAAAAA AAGTTTCCTA 1980
5 GGGAAGTATC TCTGAGGGAA CAGGCAATCT GAAGGAAGTG ACTATATTCT CCATGGCTAA 2040
GTCCATTAGG CCAAAAGNCT GGGTGGGTAT TGGTTGTICAN GCTGTCTATT GGCATATTAA 2100
AAACGTAGGC CGGANGGAAT AATTAGGTG TNATGCCGGC GGG 2143

(2) INFORMATION FOR SEQ ID NO: 229:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1025 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

CCTGCCCCAC ATTGCTTCAT TGGCCTGGCC ATGCGCCTGT ACTATGGCAG CCGCTAGTCC 60
25 CTGACAACTT CCACCTGAT TCGGACCCT GTAGATTGGG CGCCACCACC AGATCCCCCT 120
CCCAGGCCTT CCTCCCTCTC CCATCAGCAG CCCTGTAACA AGTGCCTTGT GAGAAAAGCT 180
30 GGAGAAGTGA GGCAGCCAG GTTATTCTCT GGAGGTGGT GGATGAAGGG GTACCCTAGG 240
AGATGTGAAG TGTGGGTTTG GTTAAGGAAA TGCTTACCAT CCCCCACCC CAACCAAGTT 300
CTTCAGACT AAAGAATTAA GGTAACTCA ATACCTAGGC CTGAGAAATA ACCCATCCT 360
35 TGTGGGCAG CTCCCTGCTT TGTCTGCTAT GAACAGAGTT GATGAAAGTG GGGTGTGGGC 420
AACAAAGTGC TTTCCTTGCC TACTTTAGTC ACCCAGCAGA GCCACTGGAG CTGGCTAGTC 480
40 CAGCCAGCC ATGGTGCATG ACTCTTCCAT AAGGGATCCT CACCCTTCCA CTTTCATGCA 540
AGAAGGCCCA GTTGCCACAG ATTATACAAC CATTACCCAA ACCACTCTGA CAGTCTCTC 600
CAGTTCAGC AATGCCTAGA GACATGCTCC CTGCCCTCTC CACAGTGTG CTCCCCACAC 660
45 CTAGCCTTTG TTCTGGAAAC CCCAGAGAGG GCTGGGCTTG ACTCATCTCA GGGAAATGTAG 720
CCCCTGGGCC CTGGCTTAAG CCGACACTCC TGACCTCTCT GTTCACCCCTG AGGGCTGTCT 780
50 TGAAGCCCGC TACCACTCTT GAGGCTCCTA GGAGGTACCA TGCTTCCAC TCTGGGGCCT 840
GCCCCGCTT AGCAGTCTCC CAGCTCCCAA CAGCCTGGG AAGCTCTGCA CAGAGTGACC 900
TGAGACCAGG TACAGGAAAC CTGTAGCTCA ATCAGTGTCT CTTTAACTGC ATAAGCAATA 960
55 AGATCTTAAT AAAGTCTTCT AGGCTGTAGG GTGGTTCTTA CAACCACAGC CAAAAAATA 1020
AAAAA 1025

60

(2) INFORMATION FOR SEQ ID NO: 230:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1250 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 230:

5 G C C C A C G C G T C G C C C A C G C G T C G G G G T G C G G A G T A T G G G C G C T G A T G G C C A T G G A G 60

15 G G C T A C T G G C G C T T C C T G G C G C Y G C T G G G G T C G G C A C T G C T C G T C G G C T T C C T G T C G G T G 120

A T S T T C G C C C T G T C T G G G T C C T C C A C T A C C G A G A G G G G C T T G G C T G G G A T G G G A G C G C A 180

C T A G A G T T T A A C T G G C A C C C A G T G C T S A T G G T C A C G G G C T T G T C T T C A T C C A G G C A T C 240

20 G C A T C A T C G T C T A C A G A C T G C C G T G G A C C T G G A A T G C A G C A A G C T C C T G A T G A A T C C A 300

T C C A T G C A G G G T A A A T G C A G T T G C T G C C A T T C T T G C A A T T A T C T C T G T G T G G G C C G T G T 360

25 T T G A G A A C C A C A T G T T A A C A A T A T A G C C A A T A T G T A C A G T C T G C A C A G C T G G G T T G G A C 420

T G A T A G C T G T C A T A T G C T A T T T G T T A C A G C T T C T T T C A G G T T T T C A G T C T T T C T G C T T C 480

C A T G G G C T C C G C T T C T C T C C G A G C A T T T C T C A T G C C C A T A C A T G T T T A T T C T G G A A T T G 540

30 T C A T C T T T G G A A C A G T G A T T G C A A C A G C A C T T A T G G G A T T G A C A G A G A A A C T G A T T T T T T 600

C C C T G A G A G A T C C T G C A T A C A G T A C A T T C C C G C C A G A A G G T G T T T T C G T A A A T A C G C T T G 660

35 G C C T T C T G A T C C T G G T G T T C G G G C C C T C A T T T T T G G A T A G T C A C C A G A C G C A A T G G A 720

A A C G T C C T A A G G A G C C A A A T T C T A C C A T T C T T C A T C C A A A T G G A G G C A C T G A A C A G G G A G 780

C A A G A G G T T C C A T G C C A G C C T A C T C T G G C A A C A A C A T G G A C A A T C A G A T T C A G A G T T A A 840

40 A C A R T G A A G T A G C A G C A A G G A A A G A A A C T T A G C T C T G G A T G A G G C T G G G C A G A G A T C T A 900

C C A T G T A A A A T G T T G T A G A G A T A G A G C C A T A T A A C G T C A C G T T C A A A A C T A G C T C T A C A 960

45 G T T T T G C T T C T C T A T T A G C C A T A T G A T A A T T G G G C T A T G T A G T A T C A A T A T T A C T T T A 1020

A T C A C A A A G G A T G G T T T C T T G A A T A A T T T G T A T T G A T T G A G G C C T A T G A A C T G A C C T G A 1080

A T T G G A A A G G A T G T G A T T A A T A T A A A T A A T A G C A G A T A T A A T T G T G G T T A T G T T A C C T T 1140

50 T A T C T T G T T G A G G A C C A C A A C A T T A G C A C G G T G C C T T G T G C A K A A T A G A T A C T C A A T A T G 1200

T G A A T A T G T G T C T A C T A G T A G T A A T T G G A T A A C T G G C A G C A T C C C T G A 1250

(2) INFORMATION FOR SEQ ID NO: 231:

(i) SEQUENCE CHARACTERISTICS:

484

(A) LENGTH: 1811 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 231:

	CNNGCAGTAC CGGTCNGATT CCCGGGTGCA CCCACGGGTC CGCTGCATTC CAGGGCCTTT	60
10	CAGTGGCTTT CATTCGTAAG TTCCTGGATA ACATGTTCCA TGTCTTGATG GCCCAGGTTA	120
	CCASTGTCAT TATCACAACA GTGTCGTGCC TGGTCTTGA CTTCAGGCCC TCCCTGGAAT	180
	TTTTCTTGA AGCCSCATCA GTCSTYCTCT CTATATTTAT TTATAATGCC AGCAAGCCTC	240
15	AAGTTCCGA ATACGCACCT AGGCAAGAAA GGATCCGAGA TCTAAGTGGC AATCTTTGGG	300
	AGCGTCCAG TGGGGATGA GAAGAACTAG AAAGACTTAC CAAACCCAAG AGTGATGAGT	360
20	CAGATGAAGA TACTTTCTAA CTGGTACCCA CATAGTTTGC AGCTCTCTG AACCTTATTT	420
	TCACATTTTC AGTGTTTGTA ATATTTATCT TTTCACTTTG ATAAACCAGA AATGTTTCTA	480
	AATCCTAATA TTCTTTGTCAT ATATCTAGCT ACTCCCTAAA TGGTTCCATC CAAGGCTTAG	540
25	AGTACCCAAA GGCTAAGAAA TTCTAAAGAA CTGATACAGG AGTAACAATA TGAAGAAITC	600
	ATTAATATCT CAGTACTTGA TAAATCAGAA AGTTATATGT GCAGATTATT TTCCTTGGCC	660
30	TTCAAGCTTC CAAAAAATT GTAATAATCA TGTTAGCTAT AGCTTGTATA TACACATAGA	720
	GATCAATTTG CCAATATTC ACAATCATGT AGTTCTAGTT TACATGCCAA AGTCTTCCCT	780
	TTTTAACATT ATAAAGCTA GGTGTCTCT TGAATTTGA GGCCTAGAG ATAGTCATTT	840
35	TGCAAGTAAA GAGCAACGGG ACCCTTTCTA AAAACGTTGG TTGAAGGACC TAAATACCTG	900
	GCCATACCAT AGATTTGGGA TGATGTAGTC TGTGCTAAAT ATTTTGCTGA AGAAGCAGTT	960
40	TCTCAGACAC AACATCTCAG AATTTTAATT TTTAGAAAT CATGGGAAAT TGGATTTTGT	1020
	TAATAATCTT TTGATGTTTT AAACATTGGT TCCCTAGTCA CCATAGTTAC CACTTGTATT	1080
	TTAAGTCATT TAAACAAGCC ACGGTGGGGC TTTTCTCTCC TCAGTTTGAG GAGAAAAATC	1140
45	TTGATGTCAT TACTCCTGAA TTATTACATT TTGGAGAATA AGAGGGCATT TTATTTTATT	1200
	AGTTACTAAT TCAAGCTGTG ACTATTGTAT ATCTTTCCAA GAGTTGAAAT GCTGGCTTCA	1260
50	GAATCATACC AGATTGTCAG TGAAGCTGAT GCCTAGGAAC TTTTAAAGGG ATCCTTTCAA	1320
	AAGGATCACT TAGCAAACAC ATGTTGACTT TTAAGTGATG TATGAATATT AATACTCTAA	1380
	AAATAGAAAG ACCAGTAATA TATAAGTCAC TTTACAGTGC TACTTCACAC TTAAGAGTGC	1440
55	ATGGTATTTT TCATGGTATT TTGCATGCAG CCAGTTAACT CTCGTAGATA GAGAAGTCAG	1500
	GTGATAGATG ATATTAAAAA TTAGCAAACA AAAGTGACTT GCTCAGGGTC ATGCAGCTGG	1560
60	GTGATGATAG AAGAGTGGGC TTAACTGGC AGGCCTGTAT GTTTACAGAC TACCATACTG	1620

5 TAAATATGAG CTTTATGGTG TCATTCTCAG AAACCTTATAC ATTTCTGCTC TCCTTTCTCC 1680
TAAGTTTCAT GCAGATGAAT ATAAGGTAAT ATACTATTAT ATAATTCATT TGTGATATCC 1740
ACAATAATAT GACTGGCAAG AATTGGTGA AATTGTGAAT TAAAATAATT ATTAAACCTA 1800
AAAAAAAAAN N 1811

10

(2) INFORMATION FOR SEQ ID NO: 232:

15

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2271 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

CTGACCTCAT GCGGTAGAGC CTAGCAACAG CGCAGGCTCC CAGCCGAGTC CGTTATGGCC 60
25 GCTGCCGTCC CGAAGAGGAT GAGGGGGCCA GCACAAGCGA AACTGCTGCC CGGGTCCGCC 120
ATCCAAGCCC TTGTGGGGTT GGCGCGGCCG CTGGTCTTGG CGCTCCTGCT TGTGTCCGCC 180
GCTCTATCCA GTGTGTATC ACGGACTGAT TCACCGAGCC CAACCGTACT CAACTCACAT 240
30 ATTTCTACCC CAAATGTGAA TGCTTTAACA CATGAAAACC AAACCAAACC TTCTATTTCC 300
CAAATCAGCA CCACCTCCC TCCCACGACG AGTACCAAGA AAAGTGGAGG AGCATCTGTG 360
35 GTCCCTCATC CCTCGCTAC TCCTCTGTCT CAAGAGGAAG CTGATAACAA TGAAGATCCT 420
AGTATAGAGG AGGAGGATCT TCTGATGCTG AACAGTTCTC CATCCACAGC CAAAGACACT 480
CTAGACAATG GCGATTATGG AGAACCAGAC TATGACTGGA CCACGGGCC CAGGGACGAC 540
40 GACGAGTCTG ATNGACACCT TGAAGAAAA CAGGGGTAC ATGGAATG AACAGTCAGT 600
GAAATCTTTT AAGATGCCAT CCTCAAATAT AGAAGAGGAA GACAGCCATT TCTTTTTC 660
45 TCTTATTATT TTTGCTTTTT GCATTGCTGT TGTTTACATT ACATATCACA ACAAAGGAA 720
GATTTTTCTT CTGGTTCAAA GCAGGAAATG GCGTGATGCC CTTTGTCCA AACAGTGA 780
ATACCATCGC CTAGATCAGA ATGTTAATGA GGCAATGCCT TCTTTGAAGA TTACCAATGA 840
50 TTATATTTTT TAAAGCACTG TGATTGAAT TTGCTTATGT AATTTATT TTCTGACTTT 900
TTATATGATA TTGTGCAAT GTTGCCATA GGCAATGGT ACTTAAATGA GAGGTGAGTC 960
55 TCTCTTTTGC CTGGTGCTT TGAATTAAT ATGTCACAAA CGAGTATATA ATTTTTTATC 1020
TGTACTTTTA GAGCTGAGTT TAATCAGGTG TCCAAAATGT GAGTTAAACA TTACCTTATA 1080
TTTACACTGT TAGTTTTAT TGTTTTAGAT TTATTATGCT TCTTCTGGAA GTATTAGTGA 1140
60

486

TGCTACTTTT AAAAGATCCC AAACCTGTAA CTAAATCTG ACATATCTGT TACTGCTGAC 1200
 TCACATTTCAT TCTCCGCCAT TCAATACTA TTTTATATCC ACATTTT TTTTCCCAA 1260
 5 ACTGTAATGT ACAAGGATAT GTGTGATAAT GCTTTGGATT TGAGTAATAT TTTT TTTTCT 1320
 TCCAAGAAAA CTGCTTTGGA TATTTT TAGA TAATTTAAAC ATAATTTAGG ATAATGATAT 1380
 10 TGCTCAATCT GACCACAATT TTAGGTAAAA CATTAATGT GTCAAGAAAT CTGGCAACA 1440
 GAGACTCTGC AGCTTGCACT GGACATAGAT AAAATGTTAC AGAGATACTA TTTT TTTGGT 1500
 TGGAAATTAAT ATATTAAAT TAGAAGCAGA AACTGGTAAA ATGTTAAATA CATGTACAAT 1560
 15 TGCTTTTAGT TAGCAATTGA TTGTAGCATG GGTTCCTCCA AGGTTTCAAG CAATGGGCAG 1620
 AGTTTAAAT TATATCAGAT TCGTTACTT CGTTTATAT TTTACAGTAA ATTTGAATAA 1680
 ATCTTAGGG TCATTATCAC TTAAATAATA CTGTACCTAG GTCTTTCAA TTAATTTAT 1740
 20 ACCTGAATGA AGTTGTTTGT ATACATAAAG GATATTTGTG TACAATTACC TTTT TTTCCC 1800
 CACACTGTGT TTCTTTGTT TTGTTT TTA TGGCAACTGG AAAGTATTTA CTATGGGATT 1860
 25 CATTTATGTC TGTCTTCTA TCATAAGAA TTGATCAATA TGTAAATATG TGATTGAAC 1920
 CATGGTTGAC TTACAAGTGT CACTACAGCT TTTTAGAAAA CATAGCCCTA ATATATGTTA 1980
 AGCAGGACCC GGGTGAGCCA GTGGGCTTGC GCTTTATGTA GAGCTGGAAG AAGCCGTC 2040
 30 ATCTGTCTC TTGGGCGGAC AGTGTACTTT CCTAATAGG AAGGGAAGCA CAATGGAAAT 2100
 ACCCCTGAAC CGTTTATTG CAGTAATTT TTTTATATCT GAACTATTA TTTAATATTT 2160
 35 TGAATAAGAT TTTAAAAAT AAATGGCAAA GATATAATC TAAAAAAA AAAAAAAA 2220
 AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAANA N 2271

40

(2) INFORMATION FOR SEQ ID NO: 233:

45

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1338 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

CTTCGGTTC TCCGGGCAGC TGCCACTGCT GTAGCTTCTG CCACCTGCCA CGACCGGGCC 60
 TCTCCCTGGC GTTTGGTCAC CTCTGCTTCA TTCTCCACCG CGCCTATGGT CCTCTTGGGA 120
 55 GCCAGCGTGG CGNCCTGGC GGCTCCCGGG TGGTGAGAGA GCGTCCGGG AACGATGAAG 180
 GGCTCGCAGT GCTGCTGCTG TCTAGCCAC CTCTTGGCTT CCGTCTCTCT CCTGCTGTG 240
 60 CTGCCTGAAC TAAGCGGGYC CCTGGMAGTC CTGCTGCAGG CAGCCGAGGC CGCGCCAGGT 300

487

YTTGGGCCTC CTGACCCTAG ACCAGGACAT TACCGCCGCT GCCACCGGGC CCTWACCCCT 360
 GCCCAGCAGC CGGGCCGTGG TCTGGCTGAA GCTGCGGGGG CCGCGGGGCT CCGAGGGAGG 420
 5 CAATGGCAGC AACCTGTGG CCGGGCTTGA GACGGACGAT CACGGAGGGA AGGCCGGGGA 480
 ARGCTCGGTG GGTGGCGGCC TTGCTGTGAG CCCCAACCCT GCGGACAAGC CCATGACCCA 540
 10 GCGGGCCCTG ACCGTGTGA TGGTGGTGAG CGGCGCGGTG CTGGTGTACT TCGTGGTCAG 600
 GACGGTCAGG ATGAGAAGAA GAAACCGAAA GACTAGGAGA TATGGAGTTT TGGACACTAA 660
 CATAGAAAAT ATGGAATTGA CACCTTTAGA ACAGGATGAT GAGGATGATG ACAACACGTT 720
 15 GTTTGATGCC AATCATCCTC GAAGATAAGA ATGTGCCTTT TGATGAAAGA ACTTTATCTT 780
 TCTACAATGA AGAGTGAAT TTCTATGTTT AAGGAATAAG AAGCCACTAT ATCAATGTTG 840
 20 GGGGGGTATT TAAGTTACAT ATATTNNAAC AACCTTTAAT TTGCTGTTGC AATAAATACC 900
 GTATCCTTTT ATTATATCTT TATATGTATA GAAGTACTCT GTTAATGGGC TCAGAGATGT 960
 TGGGGATAAA GTATACTGTA ATAATTTATC TGTTTGAAAA TTACTATAAA ACGGTGTTTT 1020
 25 CTGRTCGGTT TTGTTTCCT GCTTACCATA TGATTGTAAA TTGTTTTATG TATTAATCAG 1080
 TTAATGCTAA TTATTTTTCG TGATGTCATA TGTTAAAGAG CTATAAATTC CAACAACCAA 1140
 30 CTGGTGTGTA AAAATAATTT AAAATYTCCT TTAAGTAAAG GTATTTCCCA TTTTGTGGG 1200
 GAAAAGAAGC CAAATTTATT ACTTTGTGTT GGGGTTTTTA AAATATTAAG AAATGTCTAA 1260
 GTTATGTTT GCAAAACAAT AAATATGATT TTAAATCTC TTAACAAAAA AAAAAAAAC 1320
 35 CCGGGGGGGG GGCCCGGN 1338

40

(2) INFORMATION FOR SEQ ID NO: 234:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 234:

50

Met Leu Ser Thr Gly Ile Glu Val Ala Arg Pro Pro Ala Thr Leu Leu
 1 5 10 15

Gly Leu Met Phe Val Leu Thr Gly Met Pro Arg Gly Leu Arg Xaa
 20 25 30

55

(2) INFORMATION FOR SEQ ID NO: 235:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 116 amino acids

60

488

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 235:

5 Met Asn Val Val Ile Val Ile Ile Leu Phe Ser Phe Asp Ser Val Gly
 1 5 10 15

Thr Met Phe Ser Cys Asn Arg Ile Pro Lys Ile Thr Val Leu Asn Lys
 20 25 30

10 Leu Lys Phe Xaa Cys Glu Val Leu Leu Arg Ile Gln Thr Ile Gln Gly
 35 40 45

15 Phe Tyr Arg Cys Thr Arg Ile Ser Arg Tyr Lys Gly Ile Phe Pro Asp
 50 55 60

Phe Cys Gln Ser Gln Cys Met Gly Cys Asn Pro Glu Ser Xaa Met Ala
 65 70 75 80

20 Val Pro Ala Leu Val Thr Pro Ile Leu Ala His Arg Lys Lys Glu Lys
 85 90 95

Gly Met Cys Leu Phe Thr Leu Ile Ile Ala Pro Thr Arg Cys Thr His
 100 105 110

25 Tyr Phe Cys Xaa
 115

30

(2) INFORMATION FOR SEQ ID NO: 236:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 103 amino acids

35

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 236:

40 Met Ser Ser Ala Lys Ile Val Arg Gln Arg Gly Ala Val Pro Thr Tyr
 1 5 10 15

Tyr Thr Thr Glu Ala Gly Glu Ile Ile Phe Leu Val Leu Asn Trp Ser
 20 25 30

45 Leu Ser Ile Leu His Ile Val Asp Val Leu Cys Ser Lys Pro Glu Lys
 35 40 45

Ser Val Thr Glu Asp Ala Ala Ser Gly Leu Ser Gln Arg Met Thr Ala
 50 55 60

50 Leu Val Trp Arg Lys Gly Pro Asp Gly Gly Ser Arg Lys Pro Ile Leu
 65 70 75 80

55 Leu Leu Phe Phe Phe Leu Pro Leu Ile Leu Cys Phe His Ser Phe Ile
 85 90 95

His Ser Ser Asn Ile Cys Xaa
 100

60

489

(2) INFORMATION FOR SEQ ID NO: 237:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

5 Met Ile Leu Phe Pro Gln Xaa Ala Leu Arg Leu Gly Xaa Trp Pro Arg
 1 5 10 15
 Thr Trp Ser Ile Leu Xaa Lys Tyr Ser Val Asn Phe Phe Ser Ala Tyr
 20 25 30
 15 Ser Pro Met Gly Ala Val Gly Thr Glu Phe
 35 40

20

(2) INFORMATION FOR SEQ ID NO: 238:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

25 Met Ile Ile Leu Leu Leu Phe Met Leu Leu Asn Asn Val Val Leu Val
 30 1 5 10 15
 Gln Glu Asp Asn Cys Gln Arg Lys Asn Thr Val Gln Glu Arg Arg Xaa
 20 25 30
 35 Trp Ser Gln Trp Xaa
 35

40

(2) INFORMATION FOR SEQ ID NO: 239:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 128 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

45 Met Ala Ala Xaa Pro Pro Gly Cys Thr Pro Pro Xaa Leu Leu Asp Ile
 50 1 5 10 15
 Ser Trp Leu Thr Glu Ser Leu Gly Ala Gly Gln Pro Val Pro Val Glu
 20 25 30
 55 Cys Arg His Arg Leu Glu Val Ala Gly Pro Arg Lys Gly Pro Leu Ser
 35 40 45
 Pro Ala Trp Met Pro Ala Tyr Ala Cys Gln Arg Pro Thr Pro Leu Thr
 50 55 60
 60 His His Asn Thr Gly Leu Ser Glu Leu Leu Glu His Gly Val Cys Glu

490

65 70 75 80

Glu Val Glu Arg Val Arg Arg Ser Glu Arg Tyr Gln Thr Met Lys Val
 85 90 95

5 Arg Arg Ala Gly Leu Gly Pro Thr Pro Gly Met Ser Cys Pro Gly Asn
 100 105 110

10 Asp Asn Thr Val His Thr Met His Gly Glu Ala Asn Arg Gly Ser Xaa
 115 120 125

15

(2) INFORMATION FOR SEQ ID NO: 240:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 67 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

25 Met Ser Ile Leu Cys Cys Pro Xaa Leu Cys Leu Phe Phe Ser Phe Cys
 1 5 10 15

Ile Ser Ser Gly Ser Cys Pro Phe Ser His Val Ser Gln Leu Ser Phe
 20 25 30

30 Ile Ala Thr Phe Ser Gln Ser Ser Pro Val Leu Leu Val Pro Ala Tyr
 35 40 45

35 Asn Thr Tyr Leu Ser Phe Leu Ala Phe Leu Asp Cys Ala Ser Leu Thr
 50 55 60

Ser Thr Xaa
 65

40

(2) INFORMATION FOR SEQ ID NO: 241:

45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 69 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

50 Met Ser Thr Phe Gln Leu Leu Leu Leu Ile Leu Ala Gln Ser Thr Tyr
 1 5 10 15

Lys Ile Lys Ser Lys Pro Leu His Met Thr Asn His Thr Leu Leu Asn
 20 25 30

55 Ser Pro Gly Leu Asn Pro Ser Ser Pro Thr Leu Asn Phe Lys Thr Gln
 35 40 45

60 Gln His Glu Ser Val Ser Tyr Ala Cys Cys His Met Arg Ser Leu His
 50 55 60

491

His Ala Phe Ala Xaa
65

5

(2) INFORMATION FOR SEQ ID NO: 242:

- 10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 44 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

15 Met Val Ser Val Val Leu Ile Phe Ser Phe Leu Ser Leu Thr Ile Ser
 1 5 10 15
 Thr Thr Ala Ser Ala Tyr Asn Gly Asn Asp Thr Gln Gly Trp Asn Asp
 20 25 30
 20 Lys Phe His Xaa Xaa Ser Val Lys Thr Gln Thr Xaa
 35 40

25

(2) INFORMATION FOR SEQ ID NO: 243:

- 30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 51 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

35 Met Ile Ser Asp Ala Gly Ala Gly Phe Gly Val Phe Leu Leu Val Pro
 1 5 10 15
 Arg Ala Gly His Cys Trp Gly Ala Gly Lys Pro Leu Pro Ser Cys Pro
 20 25 30
 40 Ser Val Ala Ser Ile Pro Ser Trp Val Leu Pro Ser Phe Leu Glu Arg
 35 40 45
 Gly Arg Xaa
 50

45

(2) INFORMATION FOR SEQ ID NO: 244:

- 50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 43 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

55

Met Val Gln Thr Ile Gln Asp Phe Leu Ser Leu Phe Ser Thr Pro Ile
 1 5 10 15
 Phe Leu Leu Leu Leu Met Phe Glu Thr Leu Ser Leu Ala Pro Ala Trp
 20 25 30

60

Leu Lys Pro Leu Arg Val Thr Ser His Ser Xaa
35 40

5

(2) INFORMATION FOR SEQ ID NO: 245:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 61 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

15 Met Ile Leu Met Pro Gly Leu Gly Thr Ser Arg Gln Arg Ser Val Pro
1 5 10 15

Phe Val Pro Thr Leu Asn Ala Ser Thr Pro Gly Ala Met Thr Gly Pro
20 25 30

20

Thr Ala Thr Leu Thr Ser Cys Gln Trp Thr Thr Ala Cys Arg Val Ser
35 40 45

25 Trp Ala Asn Gly Trp Thr Ser Leu Arg Thr Phe Arg Xaa
50 55 60

(2) INFORMATION FOR SEQ ID NO: 246:

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

Met Ser His His Ala Gln Pro Arg Phe Leu Leu Ile Thr Met Leu Leu
1 5 10 15

40 Gln Glu Ala Lys Pro Val Ser Asn Ile Pro His Leu Leu Glu Ser Trp
20 25 30

Tyr Phe Gly Xaa
35

45

(2) INFORMATION FOR SEQ ID NO: 247:

50 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 38 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

55

Met Asn Ser Leu Phe Trp Met Ile Leu Leu Pro Val Ser Gln Asp Gln
1 5 10 15

60 Val Val Glu Gly Leu Gln Gly Gly Phe Ser Gln Ile His Met Arg Ile
20 25 30

Leu Arg Lys His Leu Xaa
35

5

(2) INFORMATION FOR SEQ ID NO: 248:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 211 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

15 Met Ser Arg Ser Xaa Asp Val Thr Asn Thr Thr Phe Leu Leu Met Ala
1 5 10 15

Ala Ser Ile Tyr Leu His Asp Gln Asn Pro Asp Ala Ala Leu Arg Ala
20 20 25 30

Leu His Gln Gly Asp Ser Leu Glu Cys Thr Ala Met Thr Val Gln Ile
35 40 45

Leu Leu Lys Leu Asp Arg Leu Asp Leu Ala Arg Lys Glu Leu Lys Arg
25 50 55 60

Met Gln Asp Leu Asp Glu Asp Ala Thr Leu Thr Gln Leu Ala Thr Ala
65 70 75 80

30 Trp Val Ser Leu Ala Thr Gly Gly Glu Lys Leu Gln Asp Ala Tyr Tyr
85 90 95

Ile Phe Gln Glu Met Ala Asp Lys Cys Ser Pro Thr Leu Leu Leu Leu
100 105 110

35 Asn Gly Gln Ala Ala Cys His Met Ala Gln Gly Arg Trp Glu Ala Ala
115 120 125

Glu Gly Leu Leu Gln Glu Ala Leu Asp Lys Asp Ser Gly Tyr Pro Glu
40 130 135 140

Thr Leu Val Asn Leu Ile Val Leu Ser Gln His Leu Gly Lys Pro Pro
145 150 155 160

45 Glu Val Thr Asn Arg Tyr Leu Ser Gln Leu Lys Asp Ala His Arg Ser
165 170 175

His Pro Phe Ile Lys Glu Tyr Gln Ala Lys Glu Asn Asp Phe Asp Arg
180 185 190

50 Leu Val Leu Gln Tyr Ala Pro Ser Ala Glu Ala Gly Pro Glu Leu Ser
195 200 205

Gly Pro Xaa
55 210

60

(2) INFORMATION FOR SEQ ID NO: 249:

494

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 548 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

Met Glu Asp Ser Glu Ala Leu Gly Phe Glu His Met Gly Leu Asp Pro
 1 5 10 15
 10 Arg Leu Leu Gln Ala Val Thr Asp Leu Gly Trp Ser Arg Pro Thr Leu
 20 25 30
 Ile Gln Glu Lys Ala Ile Pro Leu Ala Leu Glu Gly Lys Asp Leu Leu
 35 40 45
 15 Ala Arg Ala Arg Thr Gly Ser Gly Lys Thr Ala Ala Tyr Ala Ile Pro
 50 55 60
 Met Leu Gln Leu Leu Leu His Arg Lys Ala Thr Gly Pro Val Val Glu
 20 65 70 75 80
 Gln Ala Val Arg Gly Leu Val Leu Val Pro Thr Lys Glu Leu Ala Arg
 85 90 95
 25 Gln Ala Gln Ser Met Ile Gln Gln Leu Ala Thr Tyr Cys Ala Arg Asp
 100 105 110
 Val Arg Val Ala Asn Val Ser Ala Ala Glu Asp Ser Val Ser Gln Arg
 115 120 125
 30 Ala Val Leu Met Glu Lys Pro Asp Val Val Val Gly Thr Pro Ser Arg
 130 135 140
 Ile Leu Ser His Leu Gln Gln Asp Ser Leu Lys Leu Arg Asp Ser Leu
 35 145 150 155 160
 Glu Leu Leu Val Val Asp Glu Ala Asp Leu Leu Phe Ser Phe Gly Phe
 165 170 175
 40 Glu Glu Glu Leu Lys Ser Leu Leu Cys His Leu Pro Arg Ile Tyr Gln
 180 185 190
 Ala Phe Leu Met Ser Ala Thr Phe Asn Glu Asp Val Gln Ala Leu Lys
 195 200 205
 45 Glu Leu Ile Leu His Asn Pro Val Thr Leu Lys Leu Gln Glu Ser Gln
 210 215 220
 Leu Pro Gly Pro Asp Gln Leu Gln Gln Phe Gln Val Val Cys Glu Thr
 50 225 230 235 240
 Glu Glu Asp Lys Phe Leu Leu Leu Tyr Ala Leu Leu Lys Leu Ser Leu
 245 250 255
 55 Ile Arg Gly Lys Ser Leu Leu Phe Val Asn Thr Leu Glu Arg Ser Tyr
 260 265 270
 Arg Leu Arg Leu Phe Leu Glu Gln Phe Ser Ile Pro Thr Cys Val Leu
 275 280 285
 60

495

Asn Gly Glu Leu Pro Leu Arg Ser Arg Cys His Ile Ile Ser Gln Phe
 290 295 300
 5 Asn Gln Gly Phe Tyr Asp Cys Val Ile Ala Thr Asp Ala Glu Val Leu
 305 310 315 320
 Gly Ala Pro Val Lys Gly Lys Arg Arg Gly Arg Gly Pro Lys Gly Asp
 325 330 335
 10 Lys Ala Ser Asp Pro Glu Ala Gly Val Ala Arg Gly Ile Asp Phe His
 340 345 350
 His Val Ser Ala Val Leu Asn Phe Asp Leu Pro Pro Thr Pro Glu Ala
 355 360 365
 15 Tyr Ile His Arg Ala Gly Arg Thr Ala Arg Ala Asn Asn Pro Gly Ile
 370 375 380
 Val Leu Thr Phe Val Leu Pro Thr Glu Gln Phe His Leu Gly Lys Ile
 20 385 390 395 400
 Glu Glu Leu Leu Ser Gly Glu Asn Arg Gly Pro Ile Leu Leu Pro Tyr
 405 410 415
 25 Gln Phe Arg Met Glu Glu Ile Glu Gly Phe Arg Tyr Arg Cys Arg Asp
 420 425 430
 Ala Met Arg Ser Val Thr Lys Gln Ala Ile Arg Glu Ala Arg Leu Lys
 435 440 445
 30 Glu Ile Lys Glu Glu Leu Leu His Ser Glu Lys Leu Lys Thr Tyr Phe
 450 455 460
 Glu Asp Asn Pro Arg Asp Leu Gln Leu Leu Arg His Asp Leu Pro Leu
 35 465 470 475 480
 His Pro Ala Val Val Lys Pro His Leu Gly His Val Pro Asp Tyr Leu
 485 490 495
 40 Val Pro Pro Ala Leu Arg Gly Leu Val Arg Pro His Lys Lys Arg Lys
 500 505 510
 Lys Leu Ser Ser Ser Cys Arg Lys Ala Lys Arg Ala Lys Ser Gln Asn
 515 520 525
 45 Pro Leu Arg Ser Phe Lys His Lys Gly Lys Lys Phe Arg Pro Thr Ala
 530 535 540
 50 Lys Pro Ser Xaa
 545

(2) INFORMATION FOR SEQ ID NO: 250:

55

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 299 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

496

Met Thr Thr Val Pro Pro Ser Pro Arg Pro Met Ser Arg Pro Ser Glu
 1 5 10 15
 5 Arg Asn Met Arg Arg Pro Arg Gly Pro Ser Pro Leu Pro Ala Ser Pro
 20 25 30
 Arg Asn Ser Thr Pro Asp Glu Pro Asp Val His Phe Ser Lys Lys Phe
 35 40 45
 10 Leu Asn Val Phe Met Ser Gly Arg Ser Arg Ser Ser Ser Ala Glu Ser
 50 55 60
 Phe Gly Leu Phe Ser Cys Ile Ile Asn Gly Glu Glu Gln Glu Gln Thr
 65 70 75 80
 15 His Arg Ala Ile Phe Arg Phe Val Pro Arg His Glu Asp Glu Leu Glu
 85 90 95
 20 Leu Glu Val Asp Asp Pro Leu Leu Val Glu Leu Gln Ala Glu Asp Tyr
 100 105 110
 Trp Tyr Glu Ala Tyr Asn Met Arg Thr Gly Ala Arg Gly Val Phe Pro
 115 120 125
 25 Ala Tyr Tyr Ala Ile Glu Val Thr Lys Glu Pro Glu His Met Ala Ala
 130 135 140
 Leu Ala Lys Asn Ser Asp Trp Val Asp Gln Phe Arg Val Lys Phe Leu
 145 150 155 160
 Gly Ser Val Gln Val Pro Tyr His Lys Gly Asn Asp Val Leu Cys Ala
 165 170 175
 35 Ala Met Gln Lys Ile Ala Thr Thr Arg Arg Leu Thr Val His Phe Asn
 180 185 190
 Pro Pro Ser Ser Cys Val Leu Glu Ile Ser Val Arg Gly Val Lys Ile
 195 200 205
 40 Gly Val Lys Ala Asp Asp Ser Gln Glu Ala Lys Gly Asn Lys Cys Ser
 210 215 220
 His Phe Phe Gln Leu Lys Asn Ile Ser Phe Cys Gly Tyr His Pro Lys
 225 230 235 240
 Asn Asn Lys Tyr Phe Gly Phe Ile Thr Lys His Pro Ala Asp His Arg
 245 250 255
 50 Phe Ala Cys His Val Phe Val Ser Glu Asp Ser Thr Lys Ala Leu Ala
 260 265 270
 Glu Ser Val Gly Arg Ala Phe Gln Gln Phe Tyr Lys Gln Phe Val Glu
 275 280 285
 55 Tyr Thr Cys Pro Thr Glu Asp Ile Tyr Leu Glu
 290 295

60

(2) INFORMATION FOR SEQ ID NO: 251:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

5 Leu Leu Tyr Leu Leu Lys Val Xaa Val Ile Phe Val Phe Ser Ser Ser
 1 5 10 15
 Lys Gly Val Thr Leu Val Ser Met Asn Leu Thr Ser Phe Phe Val Ser
 20 25 30
 15 Ser Val Leu Ala Cys Phe Ser Xaa
 35 40

20 (2) INFORMATION FOR SEQ ID NO: 252:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 594 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

25 Met Pro Ala Ser Ser Leu Glu Ser Arg Ser Phe Leu Leu Ala Lys Lys
 1 5 10 15
 30 Ser Gly Glu Asn Val Ala Lys Phe Ile Ile Asn Ser Tyr Pro Lys Tyr
 20 25 30
 Phe Gln Lys Asp Ile Ala Glu Pro His Ile Pro Cys Leu Met Pro Glu
 35 35 40 45
 Tyr Phe Glu Pro Gln Ile Lys Asp Ile Ser Glu Ala Ala Leu Lys Glu
 50 55 60
 40 Arg Ile Glu Leu Arg Lys Val Lys Ala Ser Val Asp Met Phe Asp Gln
 65 70 75 80
 Leu Leu Gln Ala Gly Thr Thr Val Ser Leu Glu Thr Thr Asn Ser Leu
 85 90 95
 45 Leu Asp Xaa Leu Cys Tyr Tyr Gly Asp Gln Glu Pro Ser Thr Asp Tyr
 100 105 110
 His Phe Gln Gln Thr Gly Gln Ser Glu Ala Leu Glu Glu Glu Asn Asp
 50 115 120 125
 Glu Thr Ser Arg Arg Lys Ala Gly His Gln Phe Gly Val Thr Trp Arg
 130 135 140
 55 Ala Lys Asn Asn Ala Glu Arg Ile Phe Ser Leu Met Pro Glu Lys Asn
 145 150 155 160
 Glu His Ser Tyr Cys Thr Met Ile Arg Gly Met Val Lys His Arg Ala
 165 170 175

60

498

Tyr Glu Gln Ala Leu Asn Leu Tyr Thr Glu Leu Leu Asn Asn Arg Leu
 180 185 190
 5 His Ala Asp Val Tyr Thr Phe Asn Ala Leu Ile Glu Ala Thr Val Cys
 195 200 205
 Ala Ile Asn Glu Lys Phe Glu Glu Lys Trp Ser Lys Ile Leu Glu Leu
 210 215 220
 10 Leu Arg His Met Val Ala Gln Lys Val Lys Pro Asn Leu Gln Thr Phe
 225 230 235 240
 Asn Thr Ile Leu Lys Cys Leu Arg Arg Phe His Val Phe Ala Arg Ser
 245 250 255
 15 Pro Ala Leu Gln Val Leu Arg Glu Met Lys Ala Ile Gly Ile Glu Pro
 260 265 270
 Ser Leu Ala Thr Tyr His His Ile Ile Arg Leu Phe Asp Gln Pro Gly
 275 280 285
 20 Asp Pro Leu Lys Arg Ser Ser Phe Ile Ile Tyr Asp Ile Met Asn Glu
 290 295 300
 Leu Met Gly Lys Arg Phe Ser Pro Lys Asp Pro Asp Asp Asp Lys Phe
 305 310 315 320
 Phe Gln Ser Ala Met Ser Ile Cys Ser Ser Leu Arg Asp Leu Glu Leu
 325 330 335
 30 Ala Tyr Gln Val His Gly Leu Leu Lys Thr Gly Asp Asn Trp Lys Phe
 340 345 350
 Ile Gly Pro Asp Gln His Arg Asn Phe Tyr Tyr Ser Lys Phe Phe Asp
 355 360 365
 35 Leu Ile Cys Leu Met Glu Gln Ile Asp Val Thr Leu Lys Trp Tyr Glu
 370 375 380
 Asp Leu Ile Pro Ser Ala Tyr Phe Pro His Ser Gln Thr Met Ile His
 385 390 395 400
 Leu Leu Gln Ala Leu Asp Val Ala Asn Arg Leu Glu Val Ile Pro Lys
 405 410 415
 45 Ile Trp Lys Asp Ser Lys Glu Tyr Gly His Thr Phe Arg Ser Asp Leu
 420 425 430
 Arg Glu Glu Ile Leu Met Leu Met Ala Arg Asp Lys His Pro Pro Glu
 435 440 445
 Leu Gln Val Ala Phe Ala Asp Cys Ala Ala Asp Ile Lys Ser Ala Tyr
 450 455 460
 55 Glu Ser Gln Pro Ile Arg Gln Thr Ala Gln Asp Trp Pro Ala Thr Ser
 465 470 475 480
 Leu Asn Cys Ile Ala Ile Leu Phe Leu Arg Ala Gly Arg Thr Gln Glu
 485 490 495
 60

499

Ala Trp Lys Met Leu Gly Leu Phe Arg Lys His Asn Lys Ile Pro Arg
500 505 510

Ser Glu Leu Leu Asn Glu Leu Met Asp Ser Ala Lys Val Ser Asn Ser
5 515 520 525

Pro Ser Gln Ala Ile Glu Val Val Glu Leu Ala Ser Ala Phe Ser Leu
530 535 540

10 Pro Ile Cys Glu Gly Leu Thr Gln Arg Val Met Ser Asp Phe Ala Ile
545 550 555 560

Asn Gln Glu Gln Lys Glu Ala Leu Ser Asn Leu Thr Ala Leu Thr Ser
565 570 575

15 Asp Ser Asp Thr Asp Ser Ser Ser Asp Ser Asp Ser Asp Thr Ser Glu
580 585 590

Gly Lys

20

(2) INFORMATION FOR SEQ ID NO: 253:

25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 131 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 253:

Met Lys Leu Asn Leu Cys Ile Pro Asn Trp Ala Arg Cys Pro Leu Leu
1 5 10 15

35 Leu Leu Phe Pro Gln Leu Leu Pro Phe Gln Gly Glu Asp Asp Asp Pro
20 25 30

Leu Lys Ala Lys Ala Ala Asn Leu Val Glu Ala Val Pro Trp Gly Ile
35 40 45

40 Lys Ala Pro Ser Phe Gln Val Thr Cys Leu Val Arg Val Gln Leu Gln
50 55 60

Ser Cys Thr Pro Ser Arg Pro Ser Thr Leu Leu Ala Thr Ser Gln Ser
45 65 70 75 80

Pro Gly Arg Ile Ser Cys Tyr Ser Pro Leu Ser His Leu Pro Pro Val
85 90 95

50 Thr Thr Ser Ile Gln Pro Ser Pro Val Met Val Pro Phe Gln Tyr Gln
100 105 110

Ala Phe Leu Leu Gln Val Lys Glu Pro Ala Ala Gln Thr Leu Leu Gly
115 120 125

55 Gln Gln Xaa
130

60

500

(2) INFORMATION FOR SEQ ID NO: 254:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

Met Arg Tyr His Ala Gln Leu Ile Phe Cys Ile Phe Cys Xaa Phe Val
 1 5 10 15
 Phe Val Xaa Lys Xaa
 20

(2) INFORMATION FOR SEQ ID NO: 255:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 255:

Met Asn Asp Asn Ser Pro Asn His Ser Ser Ser Tyr Leu Pro Leu Pro
 1 5 10 15
 Leu Thr Ile Val Ile Leu Gln Thr Gly His Lys Gly Thr Leu Xaa
 20 25 30

(2) INFORMATION FOR SEQ ID NO: 256:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 219 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 256:

Met His Phe Leu Phe Arg Phe Ile Val Phe Phe Tyr Leu Trp Gly Leu
 1 5 10 15
 Phe Thr Ala Gln Arg Gln Lys Lys Glu Glu Ser Thr Glu Glu Val Lys
 20 25 30
 Ile Glu Val Leu His Arg Pro Glu Asn Cys Ser Lys Thr Ser Lys Lys
 35 40 45
 Gly Asp Leu Leu Asn Ala His Tyr Asp Gly Tyr Leu Ala Lys Asp Gly
 50 55 60
 Ser Lys Phe Tyr Cys Ser Arg Thr Gln Asn Glu Gly His Pro Lys Trp
 65 70 75 80
 Phe Val Leu Gly Val Gly Gln Val Ile Lys Gly Leu Asp Ile Ala Met
 85 90 95
 Thr Asp Met Cys Pro Gly Glu Lys Arg Lys Val Val Ile Pro Pro Ser
 100 105 110

501

Phe Ala Tyr Gly Lys Glu Gly Tyr Ala Glu Gly Lys Ile Pro Pro Asp
 115 120 125
 5 Ala Thr Leu Ile Phe Glu Ile Glu Leu Tyr Ala Val Thr Lys Gly Pro
 130 135 140
 Arg Ser Ile Glu Thr Phe Lys Gln Ile Asp Met Asp Asn Asp Arg Gln
 145 150 155 160
 10 Leu Ser Lys Ala Glu Ile Asn Leu Tyr Leu Gln Arg Glu Phe Glu Lys
 165 170 175
 Asp Glu Lys Pro Arg Asp Lys Ser Tyr Gln Asp Ala Val Leu Glu Asp
 15 180 185 190
 Ile Phe Lys Lys Asn Asp His Asp Gly Asp Gly Phe Ile Ser Pro Lys
 195 200 205
 20 Glu Tyr Asn Val Tyr Gln His Asp Glu Leu Xaa
 210 215

25 (2) INFORMATION FOR SEQ ID NO: 257:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 50 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 257:

Met Trp Val Ile Arg Val Phe Gln Lys Thr Phe Leu Phe Phe Val Leu
 1 5 10 15
 35 Phe Trp Ser Val His Cys Ile Ser Asp Lys Phe Gly Cys Leu Trp His
 20 25 30
 Val Cys Met Lys Arg Glu Gly Asp Xaa Asn Cys Leu Ser Phe Ser Xaa
 40 35 40 45
 Leu Xaa
 50

45

(2) INFORMATION FOR SEQ ID NO: 258:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 122 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 258:

55 Met Pro Ser Gln Thr Glu Xaa Phe Ala Ala Cys Gly Gly His Ser Leu
 1 5 10 15
 Leu Leu Val Xaa Leu Pro Leu Gly Leu Pro Phe Cys Pro Arg Ala Ala.
 20 25 30
 60

502

Leu Cys Asp Leu Pro Phe Ser Leu Pro Ser Phe Pro Gly Gln Ala Arg
 35 40 45
 5 Arg Gly Gly Ala Glu Lys Gln Gly Ala Glu Gly Arg Gly Leu Gln Val
 50 55 60
 Lys Pro Arg Gly Gln Arg Thr Phe Gln Val Ser Arg Thr Ala Pro Ala
 65 70 75 80
 10 Ala Pro Arg Ser Arg Gln Pro Arg Pro Pro Ala Ala Leu Pro Ala Leu
 85 90 95
 Gly Phe Gly Gly Arg Gly Val Ala Lys Gly Arg Phe Leu Cys Phe Trp
 100 105 110
 15 Cys Leu Tyr Met Leu Arg Ile Asp Gln Xaa
 115 120

20

(2) INFORMATION FOR SEQ ID NO: 259:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 88 amino acids
 25 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 259:

30 Met Thr Ala Phe Cys Ser Leu Leu Leu Gln Ala Gln Ser Leu Leu Pro
 1 5 10 15
 Arg Thr Met Ala Ala Pro Gln Asp Ser Leu Arg Pro Gly Glu Glu Asp
 20 25 30
 35 Glu Gly Met Gln Leu Leu Gln Thr Lys Asp Ser Met Ala Lys Gly Ala
 35 40 45
 Arg Pro Gly Ala Xaa Arg Gly Arg Ala Arg Trp Gly Leu Ala Tyr Thr
 50 55 60
 40 Leu Leu His Asn Pro Thr Leu Gln Val Phe Arg Lys Thr Ala Leu Leu
 65 70 75 80
 Gly Ala Asn Gly Ala Gln Pro Xaa
 45 85

50

(2) INFORMATION FOR SEQ ID NO: 260:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 amino acids
 55 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 260:

60 Met Ile Gln Val Ser Val Pro Leu Leu Thr Ile Met Ile Phe Leu Leu
 1 5 10 15
 Tyr Leu Gln Ile Gly Pro Gly Lys Leu Xaa

20

25

5 (2) INFORMATION FOR SEQ ID NO: 261:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acids

(B) TYPE: amino acid

10 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 261:

Met Leu Leu Asp Pro Phe Ile Leu Leu Phe Cys Leu Phe Ser Thr Ala
 1 5 10 15
 Ala Gln Ser Cys Leu Glu Phe Ile Tyr Ile Gln Phe Xaa
 20 25

20

(2) INFORMATION FOR SEQ ID NO: 262:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 44 amino acids

25 (B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

Met Lys Phe Leu Ser Ile Leu Leu Asp Asp Asn Asn Phe Xaa Leu Met
 1 5 10 15
 Leu Met Leu Ala Pro Phe Gly Cys Leu Ala Phe Glu Arg Ser Met Lys
 20 25 30
 Met Arg Asn Gly Ala Leu Gly Leu Glu Glu Val Xaa
 35 40

40 (2) INFORMATION FOR SEQ ID NO: 263:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 363 amino acids

45 (B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

Met Arg Thr Leu Phe Asn Leu Leu Trp Leu Ala Leu Ala Cys Ser Pro
 1 5 10 15
 Val His Thr Thr Leu Ser Lys Ser Asp Ala Lys Lys Ala Ala Ser Lys
 20 25 30
 Thr Leu Leu Glu Lys Ser Gln Phe Ser Asp Lys Pro Val Gln Asp Arg
 35 40 45
 Gly Leu Val Val Thr Asp Leu Lys Ala Glu Ser Val Val Leu Glu His
 50 55 60
 Arg Ser Tyr Cys Ser Ala Lys Ala Arg Asp Arg His Phe Ala Gly Asp

504

	65		70		75		80
	Val Leu Gly Tyr	Val Thr Pro Trp Asn Ser His Gly Tyr Asp	Val Thr				
		85		90		95	
5	Lys Val Phe Gly Ser Lys Phe Thr Gln Ile Ser Pro Val Trp Leu Gln						
	100		105		110		
10	Leu Lys Arg Arg Gly Arg Glu Met Phe Glu Val Thr Gly Leu His Asp						
	115		120		125		
	Val Asp Gln Gly Trp Met Arg Ala Val Arg Lys His Ala Lys Gly Leu						
	130		135		140		
15	His Ile Val Pro Arg Leu Leu Phe Glu Asp Trp Thr Tyr Asp Asp Phe						
	145		150		155		160
	Arg Asn Val Leu Asp Ser Glu Asp Glu Ile Glu Glu Leu Ser Lys Thr						
		165		170		175	
20	Val Val Gln Val Ala Lys Asn Gln His Phe Asp Gly Phe Val Val Glu						
	180		185		190		
	Val Trp Asn Gln Leu Leu Ser Gln Lys Arg Val Thr Asp Gln Leu Gly						
25	195		200		205		
	Met Phe Thr His Lys Glu Phe Glu Gln Leu Ala Pro Val Leu Asp Gly						
	210		215		220		
30	Phe Ser Leu Met Thr Tyr Asp Tyr Ser Thr Ala His Gln Pro Gly Pro						
	225		230		235		240
	Asn Ala Pro Leu Ser Trp Val Arg Ala Cys Val Gln Val Leu Asp Pro						
		245		250		255	
35	Lys Ser Lys Trp Arg Ser Lys Ile Leu Leu Gly Leu Asn Phe Tyr Gly						
	260		265		270		
	Met Asp Tyr Ala Thr Ser Lys Asp Ala Arg Glu Pro Val Val Gly Ala						
40	275		280		285		
	Arg Tyr Ile Gln Thr Leu Lys Asp His Arg Pro Arg Met Val Trp Asp						
	290		295		300		
45	Ser Gln Xaa Ser Glu His Phe Phe Glu Tyr Lys Lys Ser Arg Ser Gly						
	305		310		315		320
	Arg His Val Val Phe Tyr Pro Thr Leu Lys Ser Leu Gln Val Arg Leu						
		325		330		335	
50	Glu Leu Ala Arg Glu Leu Gly Val Gly Val Ser Ile Trp Glu Leu Gly						
	340		345		350		
	Gln Gly Leu Asp Tyr Phe Tyr Asp Leu Leu Xaa						
55	355		360				

(2) INFORMATION FOR SEQ ID NO: 264:

60

505

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 128 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

Leu Pro Thr Lys Ile Leu Val Lys Pro Asp Arg Thr Phe Glu Ile Lys
 1 5 10 15
 10 Ile Gly Gln Pro Thr Val Ser Tyr Phe Leu Lys Ala Ala Ala Gly Ile
 20 25 30
 Glu Lys Gly Ala Arg Gln Thr Gly Lys Glu Val Ala Gly Leu Val Thr
 35 40 45
 15 Leu Lys His Val Tyr Glu Ile Ala Arg Ile Lys Ala Gln Asp Glu Ala
 50 55 60
 Phe Ala Leu Gln Asp Val Pro Leu Ser Ser Val Val Arg Ser Ile Ile
 65 70 75 80
 Gly Ser Ala Arg Ser Leu Gly Ile Arg Val Val Lys Asp Leu Ser Ser
 85 90 95
 25 Glu Glu Leu Ala Ala Phe Gln Lys Glu Arg Ala Ile Phe Leu Ala Ala
 100 105 110
 Gln Lys Glu Ala Asp Leu Ala Ala Gln Glu Glu Ala Ala Lys Lys Xaa
 115 120 125
 30

35

(2) INFORMATION FOR SEQ ID NO: 265:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 54 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

45 Met Leu Leu Gln Ile His Pro Leu Leu Pro Ser Pro Thr Ile Pro His
 1 5 10 15
 Ile Leu Leu Leu Phe Leu Tyr Pro Thr Phe Ser Ile Leu Glu His Ser
 20 25 30
 50 Cys Ser Tyr Cys Ile Glu Tyr Leu Trp Val Cys Leu Leu Phe Cys Leu
 35 40 45
 Ser Leu Trp Phe Leu Xaa
 50
 55

(2) INFORMATION FOR SEQ ID NO: 266:

60

(i) SEQUENCE CHARACTERISTICS:

506

(A) LENGTH: 29 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 266:

5 Met Cys Leu Trp Cys Cys Gly Asp Val Cys Ser Gly Leu Ser Ser Leu
 1 5 10 15

10 Leu Ser Leu Cys Val Cys Cys Val Val Leu Ala Val Cys
 20 25

(2) INFORMATION FOR SEQ ID NO: 267:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 267:

Glu Gly Leu Arg Leu Leu Leu Ser Leu Pro Ala Ala Leu Pro Arg Ser
 1 5 10 15

25 Cys Cys His Pro Arg Trp Leu Pro Val Xaa
 20 25

30

(2) INFORMATION FOR SEQ ID NO: 268:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 221 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 268:

Met Phe His Gly Ile Pro Ala Thr Pro Gly Ile Gly Ala Pro Gly Asn
 1 5 10 15

40 Lys Pro Glu Leu Tyr Glu Glu Val Lys Leu Tyr Lys Asn Ala Arg Glu
 20 25 30

45 Arg Glu Lys Tyr Asp Asn Met Ala Glu Leu Phe Ala Val Val Lys Thr
 35 40 45

Met Gln Ala Leu Glu Lys Ala Tyr Ile Lys Asp Cys Val Ser Pro Ser
 50 55 60

50 Glu Tyr Thr Ala Ala Cys Ser Arg Leu Leu Val Gln Tyr Lys Ala Ala
 65 70 75 80

Phe Arg Gln Val Gln Gly Ser Glu Ile Ser Ser Ile Asp Glu Phe Cys
 85 90 95

55 Arg Lys Phe Arg Leu Asp Cys Pro Leu Ala Met Glu Arg Ile Lys Glu
 100 105 110

60 Asp Arg Pro Ile Thr Ile Lys Asp Asp Lys Gly Asn Leu Asn Arg Cys
 115 120 125

507

Ile Ala Asp Val Val Ser Leu Phe Ile Thr Val Met Asp Lys Leu Arg
 130 135 140

5 Leu Glu Ile Arg Ala Met Asp Glu Ile Gln Pro Asp Leu Arg Glu Leu
 145 150 155 160

Met Glu Thr Met His Arg Met Ser His Leu Pro Pro Asp Phe Glu Gly
 165 170 175

10 Arg Gln Thr Val Ser Gln Trp Leu Gln Thr Leu Ser Gly Met Ser Ala
 180 185 190

15 Ser Asp Glu Leu Asp Asp Ser Gln Val Arg Gln Met Leu Phe Asp Leu
 195 200 205

Glu Ser Ala Tyr Asn Ala Phe Asn Arg Phe Leu His Ala
 210 215 220

20

(2) INFORMATION FOR SEQ ID NO: 269:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 269:

30 Met Lys Xaa
 1

35 (2) INFORMATION FOR SEQ ID NO: 270:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 49 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 270:

Met Gln Ala Pro Phe Xaa His Phe Ser Phe Arg Met Phe Ser Asn Leu
 1 5 10 15

45 Tyr Cys Phe Ser Asp Phe Gln Pro Asn Ile Ser Pro Cys Pro Leu Cys
 20 25 30

His Cys Ile Leu Pro Xaa His His Val Phe Leu Leu Leu Ala Val
 35 40 45

Xaa

55

(2) INFORMATION FOR SEQ ID NO: 271:

60 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 52 amino acids

508

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 271:

5 Met Lys Leu Val Thr Met Phe Asp Lys Leu Ser Arg Asn Arg Val Ile
 1 5 10 15
 Gln Pro Met Gly Met Ser Pro Arg Gly His Leu Thr Ser Leu Gln Asp
 20 25 30
 10 Ala Met Cys Glu Thr Met Glu Gln Gln Leu Ser Ser Asp Pro Asp Ser
 35 40 45
 15 Asp Pro Asp Xaa
 50

(2) INFORMATION FOR SEQ ID NO: 272:

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

Met Ala Val Gly Glu Ala Val Phe Val Pro Leu Gln His Pro Pro Leu
 1 5 10 15
 30 Leu His Gly Ser Pro Ile Pro Lys Leu Leu Pro Gly Pro Leu Leu Xaa
 20 25 30

35

(2) INFORMATION FOR SEQ ID NO: 273:

40

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 57 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 273:

Met Asn Gly Cys His Arg Arg Lys Arg Leu His Leu Cys Lys Thr Ile
 1 5 10 15
 50 Tyr Leu Leu Trp Phe Val Phe Ser Phe Leu Leu Ser Asn Glu Val Val
 20 25 30
 Ser Ser His Trp His Ile Leu Arg Ala Val Gln Ile Ile Cys Thr Leu
 35 40 45
 55 Phe His Arg Xaa Ile Ser Ala Phe Xaa
 50 55

60

(2) INFORMATION FOR SEQ ID NO: 274:

509

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

Met Gly Trp Val Ser Ser Pro His Val Lys Arg Arg Glu Cys Val Leu
1 5 10 15
Lys Lys Pro Phe Phe Xaa
20

(2) INFORMATION FOR SEQ ID NO: 275:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

Met Phe Asn Phe Phe Lys Asn Pro Leu Leu Thr Cys Leu Phe Ile Ser
1 5 10 15
Cys Tyr Leu Tyr Leu Ser Leu Leu Val Asn Lys Val Leu Phe Ala Glu
20 25 30
Glu Gly Leu Cys Cys Thr Tyr Cys Thr Thr Ser Asn Thr Gly Glu Gly
35 40 45
Gly Val Xaa
50

(2) INFORMATION FOR SEQ ID NO: 276:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 276:

Met Xaa
1

(2) INFORMATION FOR SEQ ID NO: 277:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 66 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 277:

Met Leu Cys Thr Ile Leu Thr Val Val Ile Ile Ala Ala Gln Thr
1 5 10 15

510

Thr Arg Thr Thr Gly Ile Pro Lys Asn Ala Pro Gly Pro Ala Pro Leu
 20 25 30
 5 Cys Ala Pro Arg Ser Pro Arg Leu Phe Leu Gln Xaa Tyr Arg Gly Pro
 35 40 45
 Asn Gly Arg Pro Ala His Pro Phe Leu Gly Pro Ser Asp Leu Asp Thr
 50 55 60
 10 Ser Xaa
 65
 15 (2) INFORMATION FOR SEQ ID NO: 278:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 257 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 278:
 20 Met Leu Gly Ala Lys Pro His Trp Leu Pro Gly Pro Leu His Ser Pro
 1 5 10 15
 Gly Leu Pro Leu Val Leu Val Leu Leu Ala Leu Gly Ala Gly Trp Ala
 20 25 30
 30 Gln Glu Gly Ser Glu Pro Val Leu Leu Glu Gly Glu Cys Leu Val Val
 35 40 45
 Cys Glu Pro Gly Arg Ala Ala Ala Gly Gly Pro Gly Gly Ala Ala Leu
 50 55 60
 35 Gly Glu Ala Pro Pro Gly Arg Val Ala Phe Xaa Ala Val Arg Ser His
 65 70 75 80
 His His Glu Pro Ala Gly Glu Thr Gly Asn Gly Thr Ser Gly Ala Ile
 40 85 90 95
 Tyr Phe Asp Gln Val Leu Val Asn Glu Gly Gly Gly Phe Asp Arg Ala
 100 105 110
 45 Ser Gly Ser Phe Val Ala Pro Val Arg Gly Val Tyr Ser Phe Arg Phe
 115 120 125
 His Val Val Lys Val Tyr Asn Arg Gln Thr Val Gln Val Ser Leu Met
 130 135 140
 50 Leu Asn Thr Trp Pro Val Ile Ser Ala Phe Ala Asn Asp Pro Asp Val
 145 150 155 160
 Thr Arg Glu Ala Ala Thr Ser Ser Val Leu Leu Pro Leu Asp Pro Gly
 55 165 170 175
 Asp Arg Val Ser Leu Arg Leu Arg Arg Gly Xaa Ser Thr Gly Trp Leu
 180 185 190
 60 Glu Ile Leu Lys Phe Leu Trp Leu Pro His Leu Pro Ser Leu Lys Asp

BEST AVAILABLE COPY

511

195 200 205

Pro Ser Leu Ser Ser Thr Arg Ile Gln Pro Leu Thr Thr Phe Phe Cys
 210 215 220

5 Pro Leu Leu Pro Xaa Lys Gln Xaa Lys Gln Xaa Xaa Xaa Ser Leu Trp
 225 230 235 240

10 Leu Leu Ser His Leu Phe Ala Trp Glu Pro Val Pro Asn Thr Gln Val
 245 250 255

Xaa

15

(2) INFORMATION FOR SEQ ID NO: 279:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 102 amino acids
 (B) TYPE: amino acid
 (C) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 279:

25 Met Ala Pro Arg Ala Leu Pro Gly Ser Ala Val Leu Ala Ala Ala Val
 1 5 10 15

Phe Val Gly Gly Ala Val Ser Ser Pro Leu Val Ala Pro Asp Asn Gly
 20 25 30

30 Ser Ser Arg Thr Leu His Ser Arg Thr Glu Thr Thr Pro Ser Pro Ser
 35 40 45

35 Asn Asp Thr Gly Asn Gly His Pro Glu Tyr Ile Ala Tyr Ala Leu Val
 50 55 60

Pro Val Phe Phe Ile Met Gly Leu Phe Gly Val Leu Ile Xaa Pro Xaa
 65 70 75 80

40 Xaa Xaa Lys Lys Lys Gly Tyr Arg Cys Thr Thr Glu Ala Glu Gln Asp
 85 90 95

Ile Glu Glu Glu Lys Gly Xaa
 100

45

(2) INFORMATION FOR SEQ ID NO: 280:

50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 280:

55

Met Pro Val Thr Leu Ser Ser Leu Gly Phe Trp Val Leu Leu Ser Leu
 1 5 10 15

60 Leu Phe Pro Trp Arg Thr Asp Gln Gly Cys Gly Pro Ala Thr Cys Tyr
 20 25 30

Xaa

5

(2) INFORMATION FOR SEQ ID NO: 281:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 43 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 281:

15

Met Val Leu Gly Leu Leu Leu Leu Xaa Phe Phe Ser Phe Ser Ser
 1 5 10 15

Ser Pro Ser Pro Ser Ser Ser Leu Leu Leu Leu Ser Ser Phe Phe Phe
 20 25 30

20

Gln Ser Leu Ala Leu Ser Pro Arg Leu Glu Xaa
 35 40

25

(2) INFORMATION FOR SEQ ID NO: 282:

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 282:

35

Glu Trp Leu Val Phe Thr Phe Leu Leu Val Phe Gly Ser Pro Leu Gly
 1 5 10 15

Lys Gly Pro Leu Xaa
 20

40

(2) INFORMATION FOR SEQ ID NO: 283:

45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 283:

50

Met Ile Arg Ala Leu Ser Leu Phe Leu Leu Ile Phe Asp Ala Ala Leu
 1 5 10 15

Phe Ser Leu Ser Val Phe Val Phe Ile Gly His Leu Leu Pro Met Pro
 20 25 30

55

Lys Gly Thr Gly Leu His Ser Cys Ala Lys His Leu Ile Lys Ser Leu
 35 40 45

60

Lys Glu Asn Val Leu Pro Leu Met Asn Tyr Pro Asp Cys Lys Leu Lys
 50 55 60

Ile Asn Ile Ser Pro Xaa
65 70

5

(2) INFORMATION FOR SEQ ID NO: 284:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 75 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 284:

15 Met Gly Lys Leu Ile Arg Leu Ser Val Met Val Met Ser Val Arg Arg
 1 5 10 15
 Leu Phe Ser Ile Tyr Trp Val Leu Ser Thr Val Pro Asp Ala Val Gly
 20 25 30
 Ser Arg Gly Gly Met Glu Glu Glu Cys Ser Arg Gly Leu Cys Cys Val
 35 40 45
 Ala Gly Gln His Lys Gln Ala Lys Gly Lys Arg Gln Ala Trp Asn Lys
 25 50 55 60
 Gly Gly Glu Tyr Gln Cys Val Thr Tyr Cys Xaa
 65 70 75

30

(2) INFORMATION FOR SEQ ID NO: 285:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 285:

40 Met Pro Ala Leu Val Thr Leu Leu Leu Leu Phe Pro Leu Leu Pro Leu
 1 5 10 15
 Met Glu Ala Ser Cys His Val Met Arg Cys Pro Met Glu Arg Pro Thr
 20 25 30
 45 Xaa

50

(2) INFORMATION FOR SEQ ID NO: 286:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 17 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 286:

60 Glu Ala Pro Trp Gly Leu Leu Lys Leu Leu Leu Leu Ala Val Phe
 1 5 10 15

Xaa

5

(2) INFORMATION FOR SEQ ID NO: 287:

- 10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 287:

15 Met Gln Gln Lys Gln Lys Lys Ala Asn Glu Lys Lys Glu Glu Pro Lys
1 5 10 15

Xaa

20

(2) INFORMATION FOR SEQ ID NO: 288:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 288:

30 Met Gln Arg Lys Val Ser Asp Phe Ile Ile His Gln Arg Leu Thr Val
1 5 10 15

35 Asn Leu Cys Val Ile Ser Phe Phe Phe Phe Leu Pro Ile Cys Ile Phe
20 25 30

Ser Leu Ala Lys Lys Xaa
35

40

(2) INFORMATION FOR SEQ ID NO: 289:

- 45 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 289:

50 Met Ala Leu Leu Ile Ser Ser Leu Ile Trp Ser Xaa
1 5 10

55 (2) INFORMATION FOR SEQ ID NO: 290:

- 60 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 290:

Met Gln Met Phe Thr Val Ser Leu Leu Leu Ser Leu Leu Leu Arg Ser
 1 5 10 15
 5 Thr Asp Gln Asn His Leu Gln Leu Leu Val Gly Arg Glu Asp His Tyr
 20 25 30
 10 Gly Gly Xaa
 35

(2) INFORMATION FOR SEQ ID NO: 291:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 291:

Met Ser Glu Ser Ala Cys Ile Leu Asn Asn Gln Lys Glu Leu Xaa
 1 5 10 15
 25

(2) INFORMATION FOR SEQ ID NO: 292:

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 44 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 292:

Met Asp Leu Asp Arg Val Lys Ala Glu Ala Thr Glu Asp Ile Thr Ser
 1 5 10 15
 35 Gly Val Leu Cys Leu Leu Phe Leu Arg Leu Pro Pro Asn Ser Cys Ile
 20 25 30
 40 Phe Pro Ser Ala Val Leu Gly Ser Thr Arg Thr Xaa
 35 40

45

(2) INFORMATION FOR SEQ ID NO: 293:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 136 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 293:

Val Val Gly Thr Gly Thr Ser Leu Ala Leu Ser Ser Leu Leu Ser Leu
 1 5 10 15
 55 Leu Leu Phe Ala Gly Met Gln Met Tyr Ser Arg Gln Leu Ala Ser Thr
 20 25 30
 60 Glu Trp Leu Thr Ile Gln Gly Gly Leu Leu Gly Ser Gly Leu Phe Val

516

35 40 45
 Phe Ser Leu Thr Ala Phe Asn Asn Leu Glu Asn Leu Val Phe Gly Lys
 50 55 60
 5 Gly Phe Gln Ala Lys Ile Phe Pro Glu Ile Leu Leu Cys Leu Leu Leu
 65 70 75 80
 10 Ala Leu Phe Ala Ser Gly Leu Ile His Arg Val Cys Val Thr Thr Cys
 85 90 95
 Phe Ile Phe Ser Met Val Gly Leu Tyr Tyr Ile Asn Lys Ile Ser Ser
 100 105 110
 15 Thr Leu Tyr Gln Ala Ala Ala Pro Val Leu Thr Pro Ala Lys Val Thr
 115 120 125
 Gly Lys Ser Lys Lys Arg Asn Xaa
 130 135
 20

(2) INFORMATION FOR SEQ ID NO: 294:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 34 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 294:

30 Met Phe Ile Phe Leu Phe Leu Cys Val Leu Ser Arg Lys Ile Gln Glu
 1 5 10 15
 35 Glu Tyr Tyr Arg Leu Phe Lys Asn Val Pro Cys Cys Phe Gly Cys Leu
 20 25 30
 Arg Xaa

40

(2) INFORMATION FOR SEQ ID NO: 295:

45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 137 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 295:

50 Met Arg Thr Pro Gly Pro Leu Pro Val Leu Leu Leu Leu Ala Gly
 1 5 10 15
 Ala Pro Ala Ala Arg Pro Thr Pro Pro Thr Cys Tyr Ser Arg Met Arg
 20 25 30
 55 Ala Leu Ser Gln Glu Ile Thr Arg Asp Phe Asn Leu Leu Gln Val Ser
 35 40 45
 60 Glu Pro Ser Glu Pro Cys Val Arg Tyr Leu Pro Arg Leu Tyr Leu Asp
 50 55 60

517

Ile His Asn Tyr Cys Val Leu Asp Lys Leu Arg Asp Phe Val Ala Ser
 65 70 75 80

5 Pro Pro Cys Trp Lys Val Ala Gln Val Asp Ser Leu Lys Asp Lys Ala
 85 90 95

Arg Lys Leu Tyr Thr Ile Met Asn Ser Phe Cys Arg Arg Asp Leu Val
 100 105 110

10 Phe Leu Leu Asp Asp Cys Asn Ala Leu Glu Tyr Pro Ile Pro Val Thr
 115 120 125

Thr Val Leu Pro Asp Arg Gln Arg Xaa
 15 130 135

20 (2) INFORMATION FOR SEQ ID NO: 296:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 58 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 296:

Met Trp Leu Leu Lys Pro Ser Ala His Ser Pro Val His Xaa Leu Val
 1 5 10 15

30 Leu Leu Phe Pro Arg Gly Trp Ser Gln Pro Gly Thr His Lys Arg Gln
 20 25 30

Ile Leu Val Asn Xaa Ala Ser Leu Pro Gly Gly Cys Leu Leu Pro Trp
 35 35 40 45

Ile Trp Ser Gly Ala Ala Leu Arg Phe Xaa
 50 55

40

(2) INFORMATION FOR SEQ ID NO: 297:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 297:

Met Ser Arg Arg Ala Glu Ala Ser Ile Phe Val Leu Pro Lys Thr Leu
 50 1 5 10 15

Leu Phe Val Leu Phe Pro Ala Phe Pro Ser Pro Ala Val Gly Cys Pro
 20 25 30

55 Val Pro Xaa
 35

60 (2) INFORMATION FOR SEQ ID NO: 298:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 78 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 298:

5 Ser Cys Tyr Ile Thr Pro Trp Ser Lys Ile Gln Ser Phe Ser Leu Ser
 1 5 10 15
 10 Leu Phe Gln Phe Ile Leu Gln Glu Val Asn Ile Thr Leu Pro Glu Asn
 20 25 30
 15 Ser Val Trp Tyr Glu Arg Tyr Lys Phe Asp Ile Pro Val Phe His Leu
 35 40 45
 Asn Gly Gln Phe Leu Met Met His Arg Val Asn Thr Ser Lys Leu Glu
 50 55 60
 20 Lys Gln Leu Leu Lys Leu Glu Gln Gln Ser Thr Gly Xaa Xaa
 65 70 75

(2) INFORMATION FOR SEQ ID NO: 299:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 95 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 299:

30 Met Phe Val Leu Phe Ser Leu Pro Lys Tyr Ala Gly Leu Arg Leu Pro
 1 5 10 15
 35 Ile Pro Gly Leu Ser Ala Leu Leu Val Phe Leu Leu Ser Leu Phe Ser
 20 25 30
 40 Arg Arg Ala Gln Val Glu Leu Thr Thr Gly Arg Glu Thr Leu Pro Lys
 35 40 45
 Asn Leu Gln Gly Tyr Phe Pro Glu Phe Gly Phe Gln Val Gln Asn Phe
 50 55 60
 45 Leu Ser Cys Lys Ile Tyr Ala Ala Ser Gln Lys Gln Pro Leu Pro Pro
 65 70 75 80
 Leu Tyr Gln Leu Arg Phe Tyr Leu Lys His Met Gly Leu Pro Xaa
 85 90 95
 50

(2) INFORMATION FOR SEQ ID NO: 300:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 44 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 300:

60

519

Met Ser Ser His Trp Thr Leu Lys Ile Leu Leu Val Pro Leu Phe Tyr
 1 5 10 15

5 Leu Ser Leu Glu Phe Pro Ser Gly Phe Val Leu Cys Leu Ala Asn Asp
 20 25 30

Leu Gly Tyr His Phe Ser Ser Arg Val Arg Ser Xaa
 35 40

10

(2) INFORMATION FOR SEQ ID NO: 301:

15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 301:

20 Met Leu Val Val Asn Ile Asn Leu Val Phe Leu Leu Phe Phe Ile Phe
 1 5 10 15

Leu Cys Tyr Leu Asp Ala Cys Ile Asn Val Phe Cys Phe Tyr Xaa
 20 25 30

25

(2) INFORMATION FOR SEQ ID NO: 302:

30 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 113 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 302:

35

Met Pro Val Leu Pro Gly Arg Thr Thr Ala Leu Leu Ser Leu Thr Leu
 1 5 10 15

40 Ala Phe Ala Val Pro Cys Ser Gly Val Glu Ala Gly Pro Cys Val Pro
 20 25 30

Arg Ser His Gly Cys Ser Ser Trp Glu Ala Ser Val Cys Val Thr Ser
 35 40 45

45 Ser Thr Pro Gly Gly Ser Trp Arg Ala Arg Ala Leu Phe Pro Ser Ala
 50 55 60

Ala Trp His Arg Xaa Ala Ala Trp Asp Ser Pro Trp Thr Gln Thr Gly
 65 70 75 80

50

Asp Phe Ala Arg Gly Ala Met Gly Gly Ala Gly Ala Leu Pro Gly Gly
 85 90 95

55 Cys Val Cys Ile Ser Gly Arg Pro Arg Ala Gln Lys Leu Pro Ala Leu
 100 105 110

Xaa

60

(2) INFORMATION FOR SEQ ID NO: 303:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 303:

10 Thr His Ile His Thr His Ile Ile Ile Cys Ser Ser Val Xaa
 1 5 10

15 (2) INFORMATION FOR SEQ ID NO: 304:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 304:

20 Met Glu Asn Phe Phe Phe Ser Phe Tyr Leu Phe Leu Ile Thr Leu Ile
 1 5 10 15

25 Pro Asn Gly Arg Thr Leu Ser Thr Thr Ala Asp His Cys Lys Ile Pro
 20 25 30

30 Cys Ile Xaa
 35

35 (2) INFORMATION FOR SEQ ID NO: 305:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

40 Met Glu Leu Trp Glu Leu Ala Leu Cys Leu Leu Val Ala Leu Ser Ala
 1 5 10 15

45 His Met Phe Thr Val Gln Leu Leu Ala Asp Leu Gly Phe Leu Phe Gly
 20 25 30

50 Gly Phe Xaa
 35

(2) INFORMATION FOR SEQ ID NO: 306:

55 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 82 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 306:

60

521

Met Gly Ala Gly Ile Leu Ala Leu Leu Leu Pro Leu Glu Ser Val Leu
 1 5 10 15
 5 Thr Cys Ser Trp Ile Ser Val Ser Thr Ser Glu Arg Gln Leu Trp Gln
 20 25 30
 Ser Ser Gln Lys Ala Thr Ile Leu Ser Leu Lys Leu Asp Ser Cys Phe
 35 40 45
 10 Cys Gly His Ser Gly Leu Lys Gly Lys Asn Glu Asp Thr Asp Ser Ser
 50 55 60
 Val Pro Ile Ile Pro Ser Lys Thr His Thr His Leu Gly Lys His Leu
 65 70 75 80
 15 Ile Xaa

20

(2) INFORMATION FOR SEQ ID NO: 307:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 72 amino acids
 (B) TYPE: amino acid
 (C) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 307:

Met Phe Tyr Phe Val Leu Phe Ile Tyr Ser Ser Ser Glu Thr Trp Ser
 1 5 10 15
 Gly Ser Val Ala Gln Asp Gly Val His Gly Val Ile Ile Gly His Cys
 20 25 30
 35 Ser Val Glu Leu Pro Gly Ser Gly Asp Pro Pro Ala Ser Ala Xaa Leu
 35 40 45
 Val Ala Gly Thr Ile Gly Thr Cys Pro Thr Met Pro Gly Phe Val Tyr
 50 55 60
 40 Phe Leu Asn Asp Val Xaa Asn Xaa
 65 70

45

(2) INFORMATION FOR SEQ ID NO: 308:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 34 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 308:

Met Asp Ser Thr Leu Arg Gln Gly Arg Xaa Leu Leu Thr Leu Val Pro
 1 5 10 15
 Ala Ser Leu Phe Ser Leu Thr Leu Gly Gly Pro Gly Pro Trp Lys Asp
 20 25 30
 60 Pro Xaa

5 (2) INFORMATION FOR SEQ ID NO: 309:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 115 amino acids

(B) TYPE: amino acid

10 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 309:

Met Gln Val Val Gly Ser Trp Pro Gly Arg Val Gly Val Val Gly Leu
 1 5 10 15

Ala Phe Ser Leu Val Ile Pro Pro Pro Ala Ile Cys Ile Ala Gly Pro
 20 25 30

Ala Pro Gly Leu Gly Gly Gly Glu Arg Gln Gln Lys Gly Leu Gly Arg
 35 40 45

Gly Gly Gly Gly Leu Arg Asn Cys Pro Gly Arg Val Gly Met Ala Ala
 50 55 60

Glu Pro Gly Ala Leu Leu Cys Leu Thr Ser Arg Asp Gly Ser Leu Leu
 65 70 75 80

Leu Ser Cys Val Arg Pro His His Val Ile Lys Pro Lys Gly Thr Ala
 85 90 95

Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Xaa Xaa
 100 105 110

Gly Gly Xaa
 115

40 (2) INFORMATION FOR SEQ ID NO: 310:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 108 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 310:

Met Asp Leu Pro Gln Phe Ile Tyr Leu Phe Ile Phe Cys Phe Cys Cys
 1 5 10 15

Leu Ala Ile Val Asn Asn Ala Ser Ile Asn Ile His Ile Gln Val Ser
 20 25 30

Met Trp Leu Tyr Val Phe Ile Ser Leu Gly Tyr Leu His Gly Ser Arg
 35 40 45

Ile Leu Gly His Asn Ile Ile Leu Cys Leu Thr Ser Gln Arg Ile Ala
 50 55 60

Lys Arg Phe Phe Ile Val Ala Ala Ser Phe Thr Phe Pro Pro Ala Met
 65 70 75 80

523

Tyr Lys Asp Phe Tyr Phe Ser Ile Ser Leu His Leu Pro Thr Leu Leu
 85 90 95

5 Phe Xaa Xaa Xaa Phe Val Phe Ser Leu Leu Pro Pro
 100 105

10 (2) INFORMATION FOR SEQ ID NO: 311:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 65 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

Met Cys Ser Pro Ser Leu Ser Ser Ser Pro Pro Pro Leu Leu Gln Val
 1 5 10 15

20 Phe Phe Phe Phe Phe Phe Ser Pro His Trp Ala Ala Lys Val Val Pro
 20 25 30

25 Gln Trp Lys Xaa Arg His Pro Gln Val Ser Ser Gln Leu Leu Leu Cys
 35 40 45

Phe Leu Arg Val Asn Cys Gln Phe Leu Phe Leu Gln Glu Ile Leu Phe
 50 55 60

30 Xaa
 65

35 (2) INFORMATION FOR SEQ ID NO: 312:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 50 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 312:

Met Cys Leu Ser Arg Trp Lys Ile Phe Tyr Thr Leu Leu Ile Leu Phe
 1 5 10 15

45 Xaa Xaa Phe Ser Ile Thr Ser Glu Xaa Glu Thr Phe Tyr Met Ile Ile
 20 25 30

50 Ile His His Asn Pro Thr Gln Ile Thr Ala Ser Cys Ser Phe Thr Phe
 35 40 45

Leu Xaa
 50

55

(2) INFORMATION FOR SEQ ID NO: 313:

60 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 293 amino acids

524

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 313:

5 Met Glu Arg Pro Asp Trp Glu Thr Ala Ile Gln Lys Pro Leu Cys Ser
 1 5 10 15
 Leu Pro Ala Gly Ser Gly Asn Ala Leu Ala Ala Ser Leu Asn His Tyr
 20 25 30
 10 Ala Gly Tyr Xaa Gln Val Thr Asn Glu Asp Leu Leu Thr Asn Cys Thr
 35 40 45
 Leu Leu Leu Cys Arg Arg Leu Leu Ser Pro Met Asn Leu Leu Ser Leu
 15 50 55 60
 His Thr Ala Ser Gly Leu Arg Leu Phe Ser Val Leu Ser Leu Ala Trp
 65 70 75 80
 20 Gly Phe Ile Ala Asp Val Asp Leu Glu Ser Glu Lys Tyr Arg Arg Leu
 85 90 95
 Gly Glu Met Arg Phe Thr Leu Gly Thr Phe Leu Arg Leu Ala Ala Leu
 100 105 110
 25 Arg Thr Tyr Arg Gly Arg Leu Ala Tyr Leu Pro Val Gly Arg Val Gly
 115 120 125
 Ser Lys Thr Pro Ala Ser Pro Val Val Val Gln Gln Gly Pro Val Asp
 30 130 135 140
 Ala His Leu Val Pro Leu Glu Glu Pro Val Pro Ser His Trp Thr Val
 145 150 155 160
 35 Val Pro Asp Glu Asp Phe Val Leu Val Leu Ala Leu Leu His Ser His
 165 170 175
 Leu Gly Ser Glu Met Phe Ala Ala Pro Met Gly Arg Cys Ala Ala Gly
 180 185 190
 40 Val Met His Leu Phe Tyr Val Arg Ala Gly Val Ser Arg Ala Met Leu
 195 200 205
 Leu Arg Leu Phe Leu Ala Met Glu Lys Gly Arg His Met Glu Tyr Glu
 45 210 215 220
 Cys Pro Tyr Leu Val Tyr Val Pro Val Val Ala Phe Arg Leu Glu Pro
 225 230 235 240
 50 Lys Asp Gly Lys Gly Val Phe Ala Val Asp Gly Glu Leu Met Val Ser
 245 250 255
 Glu Ala Val Gln Gly Gln Val His Pro Asn Tyr Phe Trp Met Val Ser
 260 265 270
 55 Gly Cys Val Glu Pro Pro Pro Ser Trp Lys Pro Gln Gln Met Pro Pro
 275 280 285
 60 Pro Glu Glu Pro Leu
 290

(2) INFORMATION FOR SEQ ID NO: 314:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 68 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

Met Pro Leu Glu Gly Phe Cys Leu Val Leu Asp Ile Gly Phe Leu Leu
 1 5 10 15

15

Val Met Leu Ile Ser Leu Ala Ser Glu Cys Phe Thr Thr Cys Leu Asp
 20 25 30

Ser Phe Ser Thr Thr Glu Pro Gly Cys Lys Phe Tyr Lys Leu Leu His
 35 40 45

20

Ser Val Ser Leu Leu Asn Ile Asn Phe Asn Val Lys Ser Leu Leu Cys
 50 55 60

Ser His Ile Xaa
 65

25

(2) INFORMATION FOR SEQ ID NO: 315:

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 105 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 315:

Met Pro Leu Gln Leu Ser Gly Gln Tyr Trp Ile Ser Leu Leu Val Phe
 1 5 10 15

40

Leu Ser Leu Gln Pro Phe Pro Gln Ala Ala Ile Pro Cys Ala Leu Thr
 20 25 30

Asp Val Gly Gly Ser Cys Val Ile Cys His Ile Leu Leu Asn Cys Leu
 35 40 45

45

Cys Ile Leu Phe Thr Leu Thr Ala Pro Ser Leu Ser His Val Leu Leu
 50 55 60

Ile Lys Met Ser Leu Ser Val Cys Tyr Glu Pro Gly Ala Asp Leu Ser
 65 70 75 80

50

Asp Arg Ala Ala Thr Gly Asn Lys Lys Leu Thr Arg Ser Thr Cys Leu
 85 90 95

55

Leu Met His Ser Asn Lys Leu Cys Xaa
 100 105

60

(2) INFORMATION FOR SEQ ID NO: 316:

526

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 71 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 316:

Met Trp Gly Cys Ser Gly Leu Gly His Arg Thr Val Ser Phe Leu Leu
 1 5 10 15

Leu Leu Pro Cys Ser Phe Pro Arg Pro Cys Xaa Leu Phe Gly Leu Ile
 20 25 30

Pro Ile Ser Arg Pro Cys Lys Val Glu Ala Pro Arg Leu Ser Val Pro
 35 40 45

Xaa Leu Ser Cys Ala Ser His Pro Tyr Cys Asn Cys Pro Met Ser Thr
 50 55 60

Ser Cys Pro Leu Pro Arg Xaa
 65 70

(2) INFORMATION FOR SEQ ID NO: 317:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 317:

Met Leu Asn Val Leu Ser Lys Val Gln Gln Leu Val Ser Xaa Leu Gly
 1 5 10 15

Leu Val Thr Phe Leu Leu Asn His Ser Ala Ala Gly Gly Ser Pro Gln
 20 25 30

His Arg Trp Leu Leu Leu Xaa
 35

(2) INFORMATION FOR SEQ ID NO: 318:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 72 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 318:

Met Lys Ala Ile Ala Arg Ala Cys Leu Leu Leu Ser Leu Leu Val Leu
 1 5 10 15

Pro His Val Val Ser Glu His Leu Phe Trp His His Asn Pro Arg His
 20 25 30

Pro Val Ile Trp Pro Phe Pro Pro Phe His Leu Ile Ser Cys Ser Val
 35 40 45

527

Ser Ala Ser Thr Trp His Leu Gly Glu Xaa Leu Leu Leu Val Pro
 50 55 60

5 Ile Ala Pro Ser Val Trp Ser Xaa
 65 70

(2) INFORMATION FOR SEQ ID NO: 319:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 62 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 319:

Met Glu Gln Gly Gly Gly Pro Arg Leu Leu Leu Ile Pro Gly Leu
 1 5 10 15

20 Leu His Asn Thr Tyr Leu Ala Arg Pro Gly Asp Phe Pro Ala Gln Gly
 20 25 30

Thr Thr Glu Asn Thr Glu Cys Gln Gly Ser Pro Ser Pro Ile Ser His
 35 40 45

25 Leu Gly Lys Val Arg Ser Leu Asp Ser Asn Thr Gln Ile Xaa
 50 55 60

(2) INFORMATION FOR SEQ ID NO: 320:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 286 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 320:

40 Met Pro Leu Leu Phe Phe Ser Val Ser Thr Leu Phe Ser Gly Ser Val
 1 5 10 15

Thr Leu Gln Gln Arg Gly Met Phe Leu Pro Trp Thr Gly Thr Gly Glu
 20 25 30

45 Gln Val Leu Ala Leu Leu Trp Pro Arg Phe Glu Leu Ile Leu Glu Met
 35 40 45

Asn Val Gln Ser Val Arg Ser Thr Asp Pro Gln Arg Leu Gly Gly Leu
 50 55 60

50 Asp Thr Arg Pro His Tyr Ile Thr Arg Arg Tyr Ala Glu Phe Ser Ser
 65 70 75 80

55 Ala Leu Val Ser Ile Asn Gln Thr Ile Pro Asn Glu Arg Thr Met Gln
 85 90 95

Leu Leu Gly Gln Leu Gln Val Glu Val Glu Asn Phe Val Leu Arg Val
 100 105 110

60 Ala Ala Glu Phe Ser Ser Arg Lys Glu Gln Leu Val Phe Leu Ile Asn

528

115 120 125

Asn Tyr Asp Met Met Leu Gly Val Leu Met Glu Arg Ala Ala Asp Asp
130 135 140

5 Ser Lys Glu Val Glu Ser Phe Gln Gln Leu Leu Asn Ala Arg Thr Gln
145 150 155 160

10 Glu Phe Ile Glu Glu Leu Leu Ser Pro Pro Phe Gly Gly Leu Val Ala
165 170 175

Phe Val Lys Glu Ala Glu Ala Leu Ile Glu Arg Gly Gln Ala Glu Arg
180 185 190

15 Leu Arg Gly Glu Glu Ala Arg Val Thr Gln Leu Ile Arg Gly Phe Gly
195 200 205

Ser Ser Trp Lys Ser Ser Val Glu Ser Leu Ser Gln Asp Val Met Arg
210 215 220

20 Ser Phe Thr Asn Phe Arg Asn Gly Thr Ser Ile Ile Gln Gly Ala Leu
225 230 235 240

25 Thr Gln Leu Ile Gln Leu Tyr His Arg Phe His Arg Val Leu Ser Gln
245 250 255

Pro Gln Leu Arg Ala Leu Pro Ala Arg Ala Glu Leu Ile Asn Ile His
260 265 270

30 His Leu Met Val Glu Leu Lys Lys His Lys Pro Asn Phe Xaa
275 280 285

35 (2) INFORMATION FOR SEQ ID NO: 321:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 55 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 321:

Met Phe Arg Ala Leu Arg Asp Leu Leu Thr His Tyr Pro Gln Gln Ile
1 5 10 15

45 Leu Leu Gln Val Leu Val Val Met Tyr Gln Val Leu Gln Val Trp Glu
20 25 30

50 Leu Pro Trp Pro Glu Leu Ile His Leu Gln Gly Ile Val Pro Thr Asp
35 40 45

Gln Leu His Leu Lys Gln Xaa
50 55

55

(2) INFORMATION FOR SEQ ID NO: 322:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 59 amino acids

60

529

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 322:

5 Asp Phe Val Pro Val Leu Val Phe Val Leu Ile Lys Ala Asn Pro Pro
 1 5 10 15
 Cys Leu Leu Ser Thr Val Gln Tyr Ile Ser Ser Phe Tyr Ala Ser Cys
 20 25 30
 10 Leu Ser Gly Glu Glu Ser Tyr Trp Trp Met Gln Phe Thr Ala Ala Val
 35 40 45
 15 Glu Phe Ile Lys Thr Ile Asp Asp Arg Lys Xaa
 50 55

(2) INFORMATION FOR SEQ ID NO: 323:

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 120 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 323:

Met His Pro Ala Arg Lys Leu Leu Ser Leu Leu Phe Leu Ile Leu Met
 1 5 10 15
 30 Gly Thr Glu Leu Thr Gln Asp Ser Ala Ala Pro Asp Ser Leu Leu Arg
 20 25 30
 Ser Ser Lys Gly Ser Thr Arg Gly Ser Leu Ala Ala Ile Val Ile Trp
 35 40 45
 35 Arg Gly Lys Ser Glu Ser Arg Ile Ala Lys Thr Pro Gly Ile Phe Arg
 50 55 60
 40 Gly Gly Gly Thr Leu Val Leu Pro Pro Thr His Thr Pro Glu Trp Leu
 65 70 75 80
 Ile Leu Pro Leu Gly Ile Thr Leu Pro Leu Gly Ala Pro Glu Thr Gly
 85 90 95
 45 Gly Gly Asp Cys Ala Ala Glu Thr Trp Lys Gly Ser Gln Arg Ala Gly
 100 105 110
 Gln Leu Cys Ala Leu Leu Ala Xaa
 115 120
 50

(2) INFORMATION FOR SEQ ID NO: 324:

55

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 44 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 324:

530

Phe Phe Leu Val Val Phe Ser Leu Ser Phe Xaa Pro Ser Val Leu Thr
 1 5 10 15

5 Ser Pro Val His Xaa Pro His Cys Cys Gln Xaa Asp Xaa Ile Leu Phe
 20 25 30

Lys Asn Thr Leu Xaa Xaa Phe Xaa Ala Lys Tyr Xaa
 35 40

10

(2) INFORMATION FOR SEQ ID NO: 325:

15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 59 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 325:

20 Met Phe Ser Arg Thr Ser Asn Phe Trp Thr Phe Phe Phe Gln Phe Leu
 1 5 10 15

Ile Phe Lys Val Phe Leu Val Leu Lys Asn Xaa Phe Thr Ser Gln Lys
 20 25 30

25

Ile Xaa Xaa Ile Xaa Xaa Glu Lys Pro Lys Lys Lys Lys Xaa Arg Gly
 35 40 45

30 Gly Arg Ala Pro Ser Pro Gln Gly Gly Pro Xaa
 50 55

35

(2) INFORMATION FOR SEQ ID NO: 326:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 326:

Met Gly Leu Leu Ile Phe Met Leu Leu Ile Gly Ile His Ser Gln Cys
 1 5 10 15

45

Ser Xaa

50

(2) INFORMATION FOR SEQ ID NO: 327:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 87 amino acids

(B) TYPE: amino acid

55

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 327:

Met Val Leu Phe Cys Phe Val Leu Phe Cys Phe Val Phe Glu Met Asp
 1 5 10 15

60

531

Ser Ser Ser Val Thr Gln Ala Gly Val Gln Trp Cys Asp Leu Gly Ser
 20 25 30

5 Leu Gln Ala Pro Pro Pro Gly Phe Ser Pro Phe Ser Cys Leu Ser Leu
 35 40 45

Pro Ser Ser Trp Asp Tyr Arg Arg Pro Pro Pro Arg Pro Ala Asn Phe
 50 55 60

10 Leu Tyr Phe Leu Val Glu Thr Gly Phe His His Val Ser Gln Asp Gly
 65 70 75 80

Leu Asp Leu Leu Thr Ser Xaa
 85

15

(2) INFORMATION FOR SEQ ID NO: 328:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 538 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 328:

Met Ser Thr Lys Lys Leu Cys Ile Val Gly Gly Ile Leu Leu Val Phe
 1 5 10 15

30 Gln Ile Ile Ala Phe Leu Val Gly Gly Leu Ile Ala Pro Gly Pro Thr
 20 25 30

Thr Ala Val Ser Tyr Met Ser Val Lys Cys Val Asp Ala Arg Lys Asn
 35 40 45

35 His His Lys Thr Lys Trp Phe Val Pro Trp Gly Pro Asn His Cys Asp
 50 55 60

Lys Ile Arg Asp Ile Glu Ala Ile Pro Arg Glu Ile Glu Ala Asn
 65 70 75 80

40 Asp Ile Val Phe Ser Val His Ile Pro Leu Pro His Met Glu Met Ser
 85 90 95

45 Pro Trp Phe Gln Phe Met Leu Phe Ile Leu Gln Leu Asp Ile Ala Phe
 100 105 110

Lys Leu Asn Asn Gln Ile Arg Glu Asn Ala Glu Val Ser Met Asp Val
 115 120 125

50 Ser Leu Ala Tyr Arg Asp Asp Ala Phe Ala Glu Trp Thr Glu Met Ala
 130 135 140

His Glu Arg Val Pro Arg Lys Leu Lys Cys Thr Phe Thr Ser Pro Lys
 145 150 155 160

55 Thr Pro Glu His Glu Gly Arg Tyr Tyr Glu Cys Asp Val Leu Pro Phe
 165 170 175

60 Met Glu Ile Gly Ser Val Ala His Lys Phe Tyr Leu Leu Asn Ile Arg
 180 185 190

532

Leu Pro Val Asn Glu Lys Lys Lys Ile Asn Val Gly Ile Gly Glu Ile
 195 200 205
 5 Lys Asp Ile Arg Leu Val Gly Ile His Gln Asn Gly Gly Phe Thr Lys
 210 215 220
 Val Trp Phe Ala Met Lys Thr Phe Leu Thr Pro Ser Ile Phe Ile Ile
 225 230 235 240
 10 Met Val Trp Tyr Trp Arg Arg Ile Thr Met Met Ser Arg Pro Pro Val
 245 250 255
 15 Leu Leu Glu Lys Val Ile Phe Ala Leu Gly Ile Ser Met Thr Phe Ile
 260 265 270
 Asn Ile Pro Val Glu Trp Phe Ser Ile Gly Phe Asp Trp Thr Trp Met
 275 280 285
 20 Leu Leu Phe Gly Asp Ile Arg Gln Gly Ile Phe Tyr Ala Met Leu Leu
 290 295 300
 Ser Phe Trp Ile Ile Phe Cys Gly Glu His Met Met Asp Gln His Glu
 305 310 315 320
 25 Arg Asn His Ile Ala Gly Tyr Trp Lys Gln Val Gly Pro Ile Ala Val
 325 330 335
 30 Gly Ser Phe Cys Leu Phe Ile Phe Asp Met Cys Glu Arg Gly Val Gln
 340 345 350
 Leu Thr Asn Pro Phe Tyr Ser Ile Trp Thr Thr Asp Ile Gly Thr Glu
 355 360 365
 35 Leu Ala Met Ala Phe Ile Ile Val Ala Gly Ile Cys Leu Cys Leu Tyr
 370 375 380
 Phe Leu Phe Leu Cys Phe Met Val Phe Gln Val Phe Arg Asn Ile Ser
 385 390 395 400
 40 Gly Lys Gln Ser Ser Leu Pro Ala Met Ser Lys Val Arg Arg Leu His
 405 410 415
 Tyr Glu Gly Leu Ile Phe Arg Phe Lys Phe Leu Met Leu Ile Thr Leu
 420 425 430
 Ala Cys Ala Ala Met Thr Val Ile Phe Phe Ile Val Ser Gln Val Thr
 435 440 445
 50 Glu Gly His Trp Lys Trp Gly Gly Val Thr Val Gln Val Asn Ser Ala
 450 455 460
 Phe Phe Thr Gly Ile Tyr Gly Met Trp Asn Leu Tyr Val Phe Ala Leu
 465 470 475 480
 55 Met Phe Leu Tyr Ala Pro Ser His Lys Asn Tyr Gly Glu Asp Gln Ser
 485 490 495
 60 Asn Gly Met Gln Leu Pro Cys Lys Ser Arg Glu Asp Cys Ala Leu Phe
 500 505 510

533

Val Ser Glu Leu Tyr Gln Glu Leu Phe Ser Ala Ser Lys Tyr Ser Phe
515 520 525

5 Ile Asn Asp Asn Ala Ala Ser Gly Ile Xaa
530 535

10 (2) INFORMATION FOR SEQ ID NO: 329:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 202 amino acids

(B) TYPE: amino acid

15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 329:

Met Gly Ile Ala Leu Ala Val Leu Gly Trp Leu Ala Val Met Leu Cys
1 5 10 15
20 Cys Ala Leu Pro Met Trp Arg Val Thr Ala Phe Ile Gly Ser Asn Ile
20 25 30
25 Val Thr Ser Gln Thr Ile Trp Glu Gly Leu Trp Met Asn Cys Val Val
35 40 45
Gln Ser Thr Gly Gln Met Gln Cys Lys Val Tyr Asp Ser Leu Leu Ala
50 55 60
30 Leu Pro Gln Asp Leu Gln Ala Ala Arg Ala Leu Val Ile Ile Ser Ile
65 70 75 80
Ile Val Ala Ala Leu Gly Val Leu Leu Ser Val Val Gly Gly Lys Cys
85 90 95
35 Thr Asn Cys Leu Glu Asp Glu Ser Ala Lys Ala Lys Thr Met Ile Val
100 105 110
40 Ala Gly Val Val Phe Leu Leu Ala Gly Leu Met Val Ile Val Pro Val
115 120 125
Ser Trp Thr Ala His Asn Ile Ile Gln Asp Phe Tyr Asn Pro Leu Val
130 135 140
45 Ala Ser Gly Gln Lys Arg Glu Met Gly Ala Ser Leu Tyr Val Gly Trp
145 150 155 160
Ala Ala Ser Gly Leu Leu Leu Leu Gly Gly Gly Leu Leu Cys Cys Asn
165 170 175
50 Cys Pro Pro Arg Thr Asp Lys Pro Tyr Ser Ala Lys Tyr Ser Ala Ala
180 185 190
55 Arg Ser Ala Ala Ala Ser Asn Tyr Val Xaa
195 200

60 (2) INFORMATION FOR SEQ ID NO: 330:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 263 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 330:

Met Ala Thr Val Thr Ala Thr Thr Lys Val Pro Glu Ile Arg Asp Val
 1 5 10 15
 10 Thr Arg Ile Glu Arg Ile Gly Ala His Ser His Ile Arg Gly Leu Gly
 20 25 30
 Leu Asp Asp Ala Leu Glu Pro Arg Gln Ala Ser Gln Gly Met Val Gly
 35 40 45
 15 Gln Leu Ala Ala Arg Arg Ala Ala Gly Val Val Leu Glu Met Ile Arg
 50 55 60
 Glu Gly Lys Ile Ala Gly Arg Ala Val Leu Ile Ala Gly Gln Pro Gly
 65 70 75 80
 Thr Gly Lys Thr Ala Ile Ala Met Gly Met Ala Gln Ala Leu Gly Pro
 85 90 95
 25 Asp Thr Pro Phe Thr Ala Ile Ala Gly Ser Glu Ile Phe Ser Leu Glu
 100 105 110
 Met Ser Lys Thr Glu Ala Leu Thr Gln Ala Phe Arg Arg Ser Ile Gly
 115 120 125
 30 Val Arg Ile Lys Glu Glu Thr Glu Ile Ile Glu Gly Glu Val Val Glu
 130 135 140
 Ile Gln Ile Asp Arg Pro Ala Thr Gly Thr Gly Ser Lys Val Gly Lys
 145 150 155 160
 35 Leu Thr Leu Lys Thr Thr Glu Met Glu Thr Ile Tyr Asp Leu Gly Thr
 165 170 175
 40 Lys Met Ile Xaa Ser Leu Thr Lys Asp Lys Val Gln Ala Gly Asp Val
 180 185 190
 Ile Thr Ile Asp Lys Ala Thr Gly Lys Ile Ser Lys Leu Gly Arg Ser
 195 200 205
 45 Phe Thr Arg Ala Arg Glu Leu Arg Arg Tyr Gly Leu Pro Asp Gln Val
 210 215 220
 Arg Ala Val Pro Arg Trp Gly Ala Pro Glu Thr Gln Gly Gly Gly Ala
 225 230 235 240
 50 His Arg Val Pro Ala Arg Asp Arg Arg His Gln Leu Ser His Pro Gly
 245 250 255
 55 Leu Pro Gly Ala Leu Leu Arg
 260

60 (2) INFORMATION FOR SEQ ID NO: 331:

535

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 260 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 331:

5 Met Leu Ala Leu Leu Gly Leu Ser Gln Ala Leu Asn Ile Leu Leu Gly
 1 5 10 15
 10 Leu Lys Gly Leu Ala Pro Ala Glu Ile Ser Ala Val Cys Glu Lys Gly
 20 25 30
 15 Asn Phe Asn Val Ala His Gly Leu Ala Trp Ser Tyr Tyr Ile Gly Tyr
 35 40 45
 Leu Arg Leu Ile Leu Pro Glu Leu Gln Ala Arg Ile Arg Thr Tyr Asn
 50 55 60
 20 Gln His Tyr Asn Asn Leu Leu Arg Gly Ala Val Ser Gln Arg Leu Tyr
 65 70 75 80
 Ile Leu Leu Pro Leu Asp Cys Gly Val Pro Asp Asn Leu Ser Met Ala
 85 90 95
 25 Asp Pro Asn Ile Arg Phe Leu Asp Lys Leu Pro Gln Gln Thr Gly Asp
 100 105 110
 30 Arg Ala Gly Ile Lys Asp Arg Val Tyr Ser Asn Ser Ile Tyr Glu Leu
 115 120 125
 Leu Glu Asn Gly Gln Arg Ala Gly Thr Cys Val Leu Glu Tyr Ala Thr
 130 135 140
 35 Pro Leu Gln Thr Leu Phe Ala Met Ser Gln Tyr Ser Gln Ala Gly Phe
 145 150 155 160
 Ser Gly Glu Asp Arg Leu Glu Gln Ala Lys Leu Phe Cys Arg Thr Leu
 165 170 175
 40 Glu Asp Ile Leu Ala Asp Ala Pro Glu Ser Gln Asn Asn Cys Arg Leu
 180 185 190
 45 Ile Ala Tyr Gln Glu Pro Ala Asp Asp Ser Ser Phe Ser Leu Ser Gln
 195 200 205
 Glu Val Leu Arg His Leu Arg Gln Glu Glu Lys Glu Glu Val Thr Val
 210 215 220
 50 Gly Ser Leu Lys Thr Ser Ala Val Pro Ser Thr Ser Thr Met Ser Gln
 225 230 235 240
 Glu Pro Glu Leu Leu Ile Ser Gly Met Glu Lys Pro Leu Pro Leu Arg
 245 250 255
 55 Thr Asp Phe Ser
 260

60

(2) INFORMATION FOR SEQ ID NO: 332:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 332:

5 Met Thr Pro Gln Lys Pro Ala Leu Ala Val Leu Leu Leu Glu Val Pro
10 1 5 10 15
Leu Leu Leu Thr Leu Ser Val Leu Lys Lys Arg Cys Leu Val Thr Cys
20 25 30
15 Glu Pro Thr Ser Arg Phe Val Ser Cys Asp Leu Pro Leu Ser Val Xaa
35 40 45

20

(2) INFORMATION FOR SEQ ID NO: 333:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 334 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 333:

30 Met Ala Ala Ala Ala Trp Leu Gln Val Leu Pro Val Ile Leu Leu Leu
1 5 10 15
35 Leu Gly Ala His Pro Ser Pro Leu Ser Phe Phe Ser Ala Gly Pro Ala
20 25 30
Thr Val Ala Ala Ala Asp Arg Ser Lys Trp His Ile Pro Ile Pro Ser
35 40 45
40 Gly Lys Asn Tyr Phe Ser Phe Gly Lys Ile Leu Phe Arg Asn Thr Thr
50 55 60
Ile Phe Leu Lys Phe Asp Gly Glu Pro Cys Asp Leu Ser Leu Asn Ile
65 70 75 80
45 Thr Trp Tyr Leu Lys Ser Ala Asp Cys Tyr Asn Glu Ile Tyr Asn Phe
85 90 95
50 Lys Ala Glu Glu Val Glu Leu Tyr Leu Glu Lys Leu Lys Glu Lys Arg
100 105 110
Gly Leu Ser Gly Lys Tyr Gln Thr Ser Ser Lys Leu Phe Gln Asn Cys
115 120 125
55 Ser Glu Leu Phe Lys Thr Gln Thr Phe Ser Gly Asp Phe Met His Arg
130 135 140
Leu Pro Leu Leu Gly Glu Lys Gln Glu Ala Lys Glu Asn Gly Thr Asn
145 150 155 160

60

537

Leu Thr Phe Ile Gly Asp Lys Thr Ala Met His Glu Pro Leu Gln Thr
 165 170 175
 5 Trp Gln Asp Ala Pro Tyr Ile Phe Ile Val His Ile Gly Ile Ser Ser
 180 185 190
 Ser Lys Glu Ser Ser Lys Glu Asn Ser Leu Ser Asn Leu Phe Thr Met
 195 200 205
 10 Thr Val Glu Val Lys Gly Pro Tyr Glu Tyr Leu Thr Leu Glu Asp Tyr
 210 215 220
 Pro Leu Met Ile Phe Phe Met Val Met Cys Ile Val Tyr Val Leu Phe
 225 230 235 240
 15 Gly Val Leu Trp Leu Ala Trp Ser Ala Cys Tyr Trp Arg Asp Leu Leu
 245 250 255
 20 Arg Ile Gln Phe Trp Ile Gly Ala Val Ile Phe Leu Gly Met Leu Glu
 260 265 270
 Lys Ala Val Phe Tyr Ala Glu Phe Gln Asn Ile Arg Tyr Lys Gly Xaa
 275 280 285
 25 Ser Val Gln Gly Ala Leu Ile Leu Ala Glu Leu Leu Ser Ala Val Lys
 290 295 300
 Arg Ser Leu Ala Arg Thr Leu Val Ile Ile Val Ser Leu Gly Tyr Gly
 305 310 315 320
 30 Ile Val Lys Pro Arg Leu Glu Ser Leu Phe Ile Arg Leu Xaa
 325 330
 35
 (2) INFORMATION FOR SEQ ID NO: 334:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 200 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 334:
 40
 45 Met Val Leu Xaa Val Val Thr Leu Gly Leu Ala Leu Phe Thr Leu Cys
 1 5 10 15
 Gly Lys Phe Lys Arg Trp Lys Leu Asn Gly Ala Phe Leu Leu Ile Thr
 20 25 30
 50 Ala Phe Leu Ser Val Leu Ile Trp Val Ala Trp Met Thr Met Tyr Leu
 35 40 45
 Phe Gly Asn Val Lys Leu Gln Gln Gly Asp Ala Trp Asn Asp Pro Thr
 50 55 60
 55 Leu Ala Ile Thr Leu Ala Ala Ser Ala Gly Ser Ser Ser Ser Thr
 65 70 75 80
 60 Pro Ser Leu Arg Ser Thr Ala Pro Phe Cys Gln Pro Cys Arg Arg Thr
 85 90 95

538

Arg Pro Thr Thr Ser Thr Arg Arg Ser Pro Gly Cys Gly Arg Arg Pro
 100 105 110

5 Ser Arg Arg Thr Cys Ser Cys Arg Gly Pro Ile Trp Arg Thr Arg Pro
 115 120 125

Ser Pro Trp Met Asn Thr Met Gln Leu Ser Glu Gln Gln Asp Phe Pro
 130 135 140

10 Thr Ala Ala Trp Glu Lys Asp Pro Val Ala Ala Trp Gly Lys Asp Pro
 145 150 155 160

Ala Leu Arg Leu Glu Ala Thr Cys Ile Ser Gln Leu Arg Trp Pro Ser
 15 165 170 175

Cys Ser Thr Val Gly Pro Ser Gln Leu Leu Arg Gln Val Thr Gln Glu
 180 185 190

20 Xaa Thr Phe Gly Glu Arg Leu Xaa
 195 200

25 (2) INFORMATION FOR SEQ ID NO: 335:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 335:

Met Leu Leu His His Gln Leu Leu Ile Val Thr Leu His Leu Val Leu
 1 5 10 15

Leu Leu Ala Thr Leu Leu Val Xaa
 20

40

(2) INFORMATION FOR SEQ ID NO: 336:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 143 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 336:

Met Thr Lys Ala Leu Leu Ile Tyr Leu Val Ser Ser Phe Leu Ala Leu
 50 1 5 10 15

Asn Gln Ala Ser Leu Ile Ser Arg Cys Asp Leu Ala Gln Val Leu Gln
 20 25 30

Leu Glu Asp Leu Asp Gly Phe Glu Gly Tyr Ser Leu Ser Asp Trp Leu
 35 40 45

Cys Leu Ala Phe Val Glu Ser Lys Phe Asn Ile Ser Lys Ile Asn Glu
 50 55 60

60

539

Asn Ala Asp Gly Ser Phe Asp Tyr Gly Leu Phe Gln Ile Asn Ser His
65 70 75 80

5 Tyr Trp Cys Asn Xaa Tyr Lys Ser Tyr Ser Glu Asn Leu Cys His Val
85 90 95

Asp Cys Gln Asp Leu Leu Asn Pro Asn Leu Leu Ala Gly Ile His Cys
100 105 110

10 Ala Lys Arg Ile Val Ser Gly Ala Arg Gly Met Asn Asn Trp Val Arg
115 120 125

Met Glu Xaa Cys Thr Val Gln Ala Gly His Ser Ser Thr Gly Xaa
130 135 140

15

(2) INFORMATION FOR SEQ ID NO: 337:

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 95 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 337:

25

Met Leu Val Ile Ala Gly Gly Ile Leu Ala Ala Leu Leu Leu Leu Ile
1 5 10 15

30

Val Val Val Leu Cys Leu Tyr Phe Lys Ile His Asn Ala Leu Lys Ala
20 25 30

Ala Lys Glu Pro Glu Ala Val Ala Val Lys Asn His Asn Pro Asp Lys
35 40 45

35

Val Trp Trp Ala Lys Asn Ser Gln Ala Lys Thr Ile Ala Thr Glu Ser
50 55 60

Cys Pro Ala Leu Gln Cys Cys Glu Gly Tyr Arg Met Cys Ala Ser Phe
65 70 75 80

40

Asp Ser Leu Pro Pro Cys Cys Cys Asp Ile Asn Glu Gly Leu Xaa
85 90 95

45

(2) INFORMATION FOR SEQ ID NO: 338:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 38 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 338:

50

Met Leu Leu Lys Ser Asn Ile Leu Met Leu Asn Leu Phe Ala Ala Asn
1 5 10 15

Val Gly Ala Asn Phe Ala Leu Thr Val Glu Lys Ile Gly Met Ile Leu
20 25 30

60

Leu Asn Val Ser Gly Xaa

35

5 (2) INFORMATION FOR SEQ ID NO: 339:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 amino acids

(B) TYPE: amino acid

10 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 339:

Met Leu Val Val Ala Phe Gly Leu Leu Val Leu Tyr Ile Leu Leu Ala
 1 5 10 15
 Ser Ser Trp Lys Arg Pro Glu Pro Gly Ile Leu Thr Asp Arg Gln Pro
 20 25 30
 Leu Leu His Asp Gly Glu Xaa
 35

25 (2) INFORMATION FOR SEQ ID NO: 340:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 71 amino acids

(B) TYPE: amino acid

30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 340:

Ser Asp Pro Leu Ala Ser Ala Ser Gln Asn Ala Gly Ile Val Ser Val
 1 5 10 15
 Gly Leu Cys Thr Arg Pro Gly Pro Gln Phe Lys Asn Ala Gln Pro Pro
 20 25 30
 Phe Pro Xaa Gln Lys Ala Pro Arg Cys Leu Trp Glu Asn Gln Pro Pro
 35 40 45
 Pro Trp Arg Lys Ala Trp Asp Leu Pro Ser His Leu Gly Arg Arg Gly
 50 55 60
 Ile Cys Gly Lys Ser Phe Xaa
 65 70

50 (2) INFORMATION FOR SEQ ID NO: 341:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 85 amino acids

(B) TYPE: amino acid

55 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 341:

Tyr Val Met Ile Phe Lys Lys Glu Phe Ala Pro Ser Asp Glu Glu Leu
 1 5 10 15
 Asp Ser Tyr Arg Arg Gly Glu Glu Trp Asp Pro Gln Lys Ala Glu Glu

541

20 25 30
 Lys Arg Asn Xaa Lys Glu Leu Ala Gln Arg Gln Xaa Gly Gly Gly Ser
 35 40 45
 5 Pro Ala Gly Ala Cys Gly Gly Glu Pro Cys Gln Arg Leu Gln Gly Gln
 50 55 60
 10 Val Gln Pro Pro His Arg Gln Gly Ser Ser Gln Arg Arg Ser Pro His
 65 70 75 80
 Ala Thr Gly Gln Xaa
 85

15

(2) INFORMATION FOR SEQ ID NO: 342:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 90 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 342:

25 Met Trp Asp Trp Asp Trp Ser Ala Pro Trp Ser Trp Pro Leu Trp Leu
 1 5 10 15
 Ser Leu Ala Leu Val Cys Leu Ser Ala Gly Ala Lys Gly His Arg Ala
 20 25 30
 30 Ser Glu Ala Gly His Ala Arg Ala Leu Thr Cys Glu Met Gly Ser Glu
 35 40 45
 35 Phe Xaa Thr Ala Xaa Gly Leu Val Leu Gly Xaa Xaa Xaa Trp Thr Xaa
 50 55 60
 Xaa Asn Gly Ser Ala Gly Pro Glu Arg Arg Gly Trp Arg Pro Ala Ala
 65 70 75 80
 40 Phe Leu Ala Val Phe Leu Leu Gly Asp Xaa
 85 90

45 (2) INFORMATION FOR SEQ ID NO: 343:

50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 48 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 343:

55 Met Phe Gly Pro Thr Phe His Ser Leu Val Leu Val Pro Pro Trp Pro
 1 5 10 15
 Asn Leu Ser Leu Leu His Phe Thr Ser Pro Val Gly Gln His Ser Ser
 20 25 30
 60 Phe Leu Pro Thr Ser Leu Arg Leu Xaa Lys Lys Lys Lys Lys Lys Lys
 35 40 45

5

(2) INFORMATION FOR SEQ ID NO: 344:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 344:

15

Met Cys Ser Lys Asn Gly Phe Leu Leu Ala Trp Ser Trp Asn Ser Pro
 1 5 10 15

20

Trp Leu Pro Gln Ala Ser Leu Ala His Gly Cys Trp Gly Arg Trp Met
 20 25 30

Ser Asp Leu Val Gly Cys Ser Arg Glu Asn Lys Cys Ala Leu Arg Asp
 35 40 45

25

His Ser Glu Arg Val Gln Gly Xaa
 50 55

30

(2) INFORMATION FOR SEQ ID NO: 345:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 222 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 345:

Ser Pro Leu Xaa Phe Cys Val Val Leu Leu Leu Gln Ala Ala Arg Gly
 1 5 10 15

40

Tyr Val Val Arg Lys Pro Ala Gln Ser Arg Leu Asp Asp Asp Pro Pro
 20 25 30

Pro Ser Thr Leu Leu Lys Asp Tyr Gln Asn Val Pro Gly Ile Glu Lys
 35 40 45

45

Val Asp Asp Val Val Lys Arg Leu Leu Ser Leu Glu Met Ala Asn Lys
 50 55 60

50

Lys Glu Met Leu Lys Ile Lys Gln Glu Gln Phe Met Lys Lys Ile Val
 65 70 75 80

Ala Asn Pro Glu Asp Thr Arg Ser Leu Glu Ala Arg Ile Ile Ala Leu
 85 90 95

55

Ser Val Lys Ile Arg Ser Tyr Glu Glu His Leu Glu Lys His Arg Lys
 100 105 110

Asp Lys Ala His Lys Arg Tyr Leu Leu Met Ser Ile Asp Gln Arg Lys
 115 120 125

60

543

Lys Met Leu Lys Asn Leu Arg Asn Thr Asn Tyr Asp Val Phe Glu Lys
 130 135 140
 Ile Cys Trp Gly Leu Gly Ile Glu Tyr Thr Phe Pro Pro Leu Tyr Tyr
 145 150 155 160
 Arg Arg Ala His Arg Arg Phe Val Thr Lys Lys Ala Leu Cys Ile Arg
 165 170 175
 Val Phe Gln Glu Thr Gln Lys Leu Lys Lys Arg Arg Arg Ala Leu Lys
 180 185 190
 Ala Ala Ala Ala Ala Gln Lys Gln Ala Lys Arg Arg Asn Pro Asp Ser
 195 200 205
 Pro Ala Lys Ala Ile Pro Lys Thr Leu Lys Asp Ser Gln Xaa
 210 215 220

20

(2) INFORMATION FOR SEQ ID NO: 346:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 64 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 346:

Met Gly Ala Pro Ala Ala Ser Leu Leu Leu Leu Leu Leu Phe Ala
 1 5 10 15
 Cys Cys Trp Ala Pro Gly Gly Ala Asn Leu Ser Gln Asp Asp Ser Gln
 20 25 30
 Pro Trp Thr Ser Asp Glu Thr Val Val Ala Gly Gly Thr Val Val Leu
 35 40 45
 Lys Cys Gln Val Lys Asp His Glu Asp Ser Ser Leu Gln Trp Ser Xaa
 50 55 60

45

(2) INFORMATION FOR SEQ ID NO: 347:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 154 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 347:

Met Val Ala Pro Val Trp Tyr Leu Val Ala Ala Ala Leu Leu Val Gly
 1 5 10 15
 Phe Ile Leu Phe Leu Thr Arg Ser Arg Gly Arg Ala Ala Ser Ala Gly
 20 25 30
 Gln Glu Pro Leu His Asn Glu Glu Leu Ala Gly Ala Gly Arg Val Ala

544

35 40 45

Gln Pro Gly Pro Leu Glu Pro Glu Glu Pro Arg Ala Gly Gly Arg Pro
50 55 60

5 Arg Arg Arg Arg Asp Leu Gly Ser Arg Leu Gln Ala Gln Arg Arg Ala
65 70 75 80

Gln Arg Val Ala Trp Ala Glu Ala Asp Glu Asn Glu Glu Glu Ala Val
85 90 95

10 Ile Leu Ala Gln Glu Glu Glu Gly Val Glu Lys Pro Ala Glu Xaa His
100 105 110

15 Leu Ser Gly Lys Ile Gly Ala Lys Lys Leu Arg Xaa Xaa Glu Glu Lys
115 120 125

Gln Ala Arg Lys Ala Gln Xaa Glu Ala Glu Glu Ala Glu Arg Glu Xaa
130 135 140

20 Arg Lys Arg Leu Glu Ser Gln Arg Glu Xaa
145 150

25

(2) INFORMATION FOR SEQ ID NO: 348:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 348:

30

Met Gln Lys Cys Met Leu Ser Ala Leu Val Phe His Ile Gln Trp Ser
1 5 10 15

35

Xaa

40

(2) INFORMATION FOR SEQ ID NO: 349:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 349:

45

Met Leu Val Cys Ser Phe Leu Phe Leu Xaa
1 5 10

50

55

(2) INFORMATION FOR SEQ ID NO: 350:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

60

545

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 350:

Val Ile Glu Leu Cys Val Ser Leu Arg Ser Leu Asn Phe Xaa
 1 5 10

(2) INFORMATION FOR SEQ ID NO: 351:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 351:

Met Cys Glu Phe Xaa Xaa Xaa Ile Met Xaa Leu Ala Gly Tyr Phe Ala
 1 5 10 15

Cys Xaa

(2) INFORMATION FOR SEQ ID NO: 352:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 62 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 352:

Met Val Gly Gly Tyr Val Ser Ser Phe Ser Phe Pro Pro Val Ser Ser
 1 5 10 15

Ser Leu Leu Leu Pro Ala Ser Phe Ala Phe Pro Phe Leu Pro Gly Thr
 20 25 30

Pro Cys Pro Phe Leu Tyr Phe Leu Pro Ser Pro Phe Ser Pro Leu Pro
 35 40 45

Leu Ser Leu Thr Arg Ser Asn Ser Phe Leu Leu Asn Gly Xaa
 50 55 60

(2) INFORMATION FOR SEQ ID NO: 353:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 353:

Glu Lys Lys Ser Met Ser Val Ser Asp Ile Tyr Ala Leu Glu Ser Leu
 1 5 10 15

Gly Arg Ser Leu Phe Thr Leu Asn Ser Met Cys Leu Pro Leu Ser Phe
 20 25 30

Xaa

5 (2) INFORMATION FOR SEQ ID NO: 354:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 245 amino acids

(B) TYPE: amino acid

10 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 354:

Met Gly Gly Ala Ser Arg Arg Val Glu Ser Gly Ala Trp Ala Tyr Leu
 1 5 10 15
 Ser Pro Leu Val Leu Arg Lys Glu Leu Glu Ser Leu Val Glu Asn Glu
 20 25 30
 Gly Ser Glu Val Leu Ala Leu Pro Glu Leu Pro Ser Ala His Pro Ile
 35 40 45
 Ile Phe Trp Asn Leu Leu Trp Tyr Phe Gln Arg Leu Arg Leu Pro Ser
 50 55 60
 Ile Leu Pro Gly Leu Val Leu Ala Ser Cys Asp Gly Pro Ser Xaa Ser
 65 70 75 80
 Gln Ala Pro Ser Pro Trp Leu Thr Pro Asp Pro Ala Ser Val Gln Val
 85 90 95
 Arg Leu Leu Trp Asp Val Leu Thr Pro Asp Pro Asn Ser Cys Pro Pro
 100 105 110
 Leu Tyr Val Leu Trp Arg Val His Ser Gln Ile Pro Gln Arg Val Val
 115 120 125
 Trp Pro Gly Pro Val Pro Ala Ser Leu Ser Leu Ala Leu Leu Glu Ser
 130 135 140
 Val Leu Arg His Val Gly Leu Asn Glu Val His Lys Ala Val Gly Leu
 145 150 155 160
 Leu Leu Glu Thr Leu Gly Pro Pro Pro Thr Gly Leu His Leu Gln Arg
 165 170 175
 Gly Ile Tyr Arg Glu Ile Leu Phe Leu Thr Met Ala Ala Leu Gly Lys
 180 185 190
 Asp His Val Asp Ile Val Ala Phe Asp Lys Lys Tyr Lys Ser Ala Phe
 195 200 205
 Asn Lys Leu Ala Ser Ser Met Gly Lys Glu Glu Leu Arg His Arg Arg
 210 215 220
 Ala Gln Met Pro Thr Pro Lys Ala Ile Asp Cys Arg Lys Cys Phe Gly
 225 230 235 240
 Ala Pro Pro Glu Cys
 245

60

(2) INFORMATION FOR SEQ ID NO: 355:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 355:

10

Met Lys Phe Ser Leu Leu Phe Leu Pro Met Leu Leu Ile Leu Lys Pro
 1 5 10 15

15

Asp Leu Phe His Ile Ser Ile Cys Thr Leu Ala Ala Cys Gly Leu Thr
 20 25 30

Phe Pro Xaa
 35

20

(2) INFORMATION FOR SEQ ID NO: 356:

25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 356:

30

Met Leu Phe Phe Phe Ile Leu His Leu Leu Ser Ile Met Ser Phe Leu
 1 5 10 15

Ser Pro Asp Ile Met Xaa
 20

35

(2) INFORMATION FOR SEQ ID NO: 357:

40

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 98 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 357:

45

Met Phe Gly Leu Leu Val Glu Ser Gln Thr Leu Leu Glu Glu Asn Ala
 1 5 10 15

50

Val Gln Gly Thr Glu Arg Thr Leu Gly Leu Asn Ile Ala Pro Phe Ile
 20 25 30

Asn Gln Phe Gln Val Pro Ile Arg Val Phe Leu Asp Leu Ser Ser Leu
 35 40 45

55

Pro Cys Ile Pro Leu Ser Lys Pro Val Glu Leu Leu Arg Leu Asp Leu
 50 55 60

60

Met Thr Pro Tyr Leu Asn Thr Ser Asn Arg Glu Val Lys Val Tyr Val
 65 70 75 80

548

Cys Xaa Ile Trp Glu Asp Leu Thr Ala Ile Pro Phe Trp Val Ser Tyr
 85 90 95

Val Pro

(2) INFORMATION FOR SEQ ID NO: 358:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 78 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 358:

Met Phe Gly Ala His Arg Xaa Trp Gln Gly Ser Val Leu Leu Phe Leu
 1 5 10 15

Ser Phe Ala Trp Gly Asn Gly Gly Ser Val Thr Phe Ser Asp Val Pro
 20 25 30

Arg Val Met Pro Leu Ala Gly Gly Pro Xaa Xaa Gln Val Ser Ser Thr
 35 40 45

Pro Arg Pro Pro Pro His Gln Val Thr Ser Ser Pro Gly Leu Glu Ser
 50 55 60

Ala His Ile Val Cys Pro Glu Arg Lys Lys Lys Lys Lys Lys
 65 70 75

(2) INFORMATION FOR SEQ ID NO: 359:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 359:

Thr Leu Leu Xaa Phe Leu Xaa Leu Leu Thr Thr Glu Gly Gly Arg Glu
 1 5 10 15

Asn Ile Phe Xaa Gly Arg Ile Leu Xaa Leu Gln Xaa Ser Pro Xaa
 20 25 30

(2) INFORMATION FOR SEQ ID NO: 360:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 57 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 360:

Met Leu Ser Phe Phe Ile Cys Leu Leu Ile Phe Val His Leu Leu Leu
 1 5 10 15